```
            GenCore version 5.1.7
     Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on:    March 16, 2006, 15:49:10 ; Search time 188 Seconds
                                       (without alignments)
                              495.470 Million cell updates/sec

Title:          US-09-989-890-238
Perfect score:  1134
Sequence:       1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212
Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5
Searched:       2443163 seqs, 439378781 residues
Total number of hits satisfying chosen parameters:    2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq_21:*
         1:   geneseqp19980s:*
         2:   geneseqp19990s:*
         3:   geneseqp20000s:*
         4:   geneseqp20010s:*
         5:   geneseqp20020s:*
         6:   geneseqp20030as:*
         7:   geneseqp20030bs:*
         8:   geneseqp20040s:*
         9:   geneseqp20050s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	1134	100.0	212	6	ABJ18429	ABJ18429 Breast sp
2	1134	100.0	217	7	ADD01203	ADD01203 Human nuc
3	1134	100.0	217	9	AEA20112	AEA20112 Novel hum
4	1069.5	94.3	213	5	ABU60984	ABU60984 Lung spec
5	1069.5	94.3	213	5	ABU60939	ABU60939 Lung spec
6	1065.5	94.0	218	5	ABJ05501	ABJ05501 Human bre
7	154	13.6	559	8	ADQ67157	ADQ67157 Novel hum
8	154	13.6	1127	4	AAB95541	AAB95541 Human pro
9	154	13.6	1480	7	ADJ70433	ADJ70433 Human hea
10	154	13.6	2427	8	ADR09758	ADR09758 Human pro
11	154	13.6	2735	6	ABR41356	ABR41356 Human DIT
12	154	13.6	2971	3	AAB41231	AAB41231 Human ORF
13	152	13.4	2971	7	ADC35084	ADC35084 Human bre
14	152	13.4	2972	4	AAB50363	AAB50363 Human SRC
15	152	13.4	3118	4	AAB50362	AAB50362 Human SRC
16	149.5	13.2	1828	4	ABB50164	ABB50164 Human tra
17	141	12.4	1008	4	ABB11527	ABB11527 Human apo
18	140.5	12.4	325	7	ABM87858	ABM87858 Rice abio
19	138.5	12.2	532	3	AAY84593	AAY84593 Amino aci
20	138	12.2	346	7	ABM89803	ABM89803 Rice abio
21	136.5	12.0	415	4	ABG30150	ABG30150 Novel hum
22	135.5	11.9	693	5	ABP69529	ABP69529 Human pol
23	135.5	11.9	842	4	ABB66631	ABB66631 Drosophil
24	135.5	11.9	864	4	ABB71319	ABB71319 Drosophil
25	135	11.9	135	7	ADM05164	ADM05164 Human pro
26	135	11.9	135	8	ADS15467	ADS15467 HSV-1 pol
27	134.5	11.9	308	7	ABO69351	ABO69351 Pseudomon
28	134.5	11.9	983	4	AAM30853	AAM30853 Peptide #
29	134.5	11.9	983	4	ABB32160	ABB32160 Peptide #
30	134.5	11.9	983	4	ABB22702	ABB22702 Protein #
31	134.5	11.9	983	4	AAM70534	AAM70534 Human bon
32	134.5	11.9	983	5	AAM58086	AAM58086 Human bra
33	134.5	11.9	983	5	ABG40187	ABG40187 Human pep
34	134.5	11.9	1482	9	ADX05726	ADX05726 Cyclin-de
35	134.5	11.9	1518	6	AAE33868	AAE33868 Human RIZ
36	134.5	11.9	1718	6	ABU62117	ABU62117 Human RIZ
37	134.5	11.9	1719	2	AAR92100	AAR92100 Human RIZ
38	134.5	11.9	1719	3	AAB12113	AAB12113 Human Rb-
39	134.5	11.9	1719	3	AAB12029	AAB12029 Human Rb-
40	134.5	11.9	1719	5	AAB12114	AAB12114 Human mut
41	134.5	11.9	1719	5	AAU10788	AAU10788 Protein B
42	134.5	11.9	1719	5	AAU80804	AAU80804 Human Rb-
43	134.5	11.9	1719	6	AAE33867	AAE33867 Human RIZ
44	134.5	11.9	1719	6	ABU62110	ABU62110 Human ret
45	134.5	11.9	1719	6	ABU62116	ABU62116 Human RIZ

ALIGNMENTS

```
RESULT 1
ABJ18429
ID   ABJ18429 standard; protein; 212 AA.
XX
AC   ABJ18429;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Breast specific related amino acid sequence SEQ ID No 238.
XX
KW   Cytostatic; BSP-agonist; BSP-antagonist; vaccine; gene therapy; cancer;
KW   metastatic; breast cancer; breast specific; human.
XX
OS   Homo sapiens.
XX
PN   WO200277232-A2.
XX
PD   03-OCT-2002.
XX
PF   21-NOV-2001; 2001WO-US043815.
XX
PR   22-NOV-2000; 2000US-0252509P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S, Macina RA, Recipon H, Pluta J, Sun Y, Liu C;
XX
DR   WPI; 2003-018927/01.
XX
PT   New isolated nucleic acid molecule, useful for treating breast cancer,
PT   and diagnosing or monitoring the presence of metastases of breast cancer
PT   in a patient.
XX
PS   Claim 11; Page 355; 377pp; English.
XX
CC   The invention relates to a novel isolated nucleic acid molecule
CC   comprising: a sequence encoding a sequence comprising 11-1518 amino acids
CC   ; a sequence comprising 190-8144 bp; or a sequence that selectively
CC   hybridises to, or having at least 60% identity with the 11-1518 amino
CC   acid or 190-8144 nt sequence. The polypeptide and the nucleic acid are
CC   useful for treating breast cancer, and diagnosing or monitoring the
CC   presence of metastases of breast cancer in a patient. The polynucleotides
CC   of the invention can be used to treat disorders by gene therapy. This
CC   sequence represents a breast specific related polypeptide of the
CC   invention
XX
SQ   Sequence 212 AA;
```

Query Match
Best Local Similarity   100.0%;   Score 1134; DB 6;   Length 212;
Matches 212; Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    1  SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA   60

Qy   61  AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP  120

Qy  121  WAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  WAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR  180

Qy  181  STILSMSRTWTCRRWAVAPCRABKLMCSSSRS  212
         ||||||||||||||||||||||||||||||||
Db  181  STILSMSRTWTCRRWAVAPCRABKLMCSSSRS  212
```

RESULT 2
ADD01203
ID   ADD01203 standard; protein; 212 AA.
XX
AC   ADD01203;
XX
DT   01-JAN-2004 (first entry)
XX
DE   Human nucleic acid-associated protein NAAP-41 SEQ ID NO:41.
XX
KW   human; nucleic acid-associated protein; NAAP; cytostatic;
KW   antiarteriosclerotic; anti-HIV; antiallergic; cerebroprotective;
KW   antiparkinsonian; anticonvulsant; nootropic; neuroprotective;
KW   antiinflammatory; ophthalmological; thyromimetic; antiarthritic;
KW   hepatotropic; antibacterial; virucide; protozoacide; antiparasitic;
KW   fungicide; gene therapy; cell proliferative disease; cancer;
KW   atherosclerosis; hepatitis; neurological disorder; Parkinson's disease;
KW   Alzheimer's disease; stroke; epilepsy; developmental disorder;
KW   renal tubular acidosis; anaemia; glaucoma; hypothyroidism;
KW   autoimmune disorder; inflammatory disorder; AIDS; allergy;
KW   atopic dermatitis; arthritis; infection.
XX
OS   Homo sapiens.
XX
PN   WO2003054219-A2.
XX
PD   03-JUL-2003.
XX
PF   18-DEC-2002; 2002WO-US041115.
XX
PR   19-DEC-2001; 2001US-0343004P.
PR   11-JAN-2002; 2002US-0347633P.
PR   25-JAN-2002; 2002US-0351749P.
PR   22-FEB-2002; 2002US-0359498P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Azimzai Y, Baughn MR, Becha SD, Borowsky ML, Burford N;
PI   Elliott VS, Emerling BM, Forsythe IJ, Gorvad AE, Griffin JA;
PI   Kable AE, Khare R, Lal PG, Lee EA, Lee SY, Li JX, Marquis JP;
PI   Ramkumar J, Richardson TW, Sprague WW, Swarnakar A, Tang YT;
PI   Chawla NK, Warren BA, Yue H;
XX
DR   WPI; 2003-559157/52.
DR   N-PSDB; ADD01260.
XX
PT   New human nucleic acid-associated proteins (NAAP), useful for diagnosing,
PT   treating and preventing diseases or conditions associated with the
PT   aberrant NAAP expression, e.g. cancer, hepatitis, AIDS, atherosclerosis,
PT   infections.
XX

PS   Claim 1; SEQ ID NO 41; 405pp; English.
XX
CC   The present invention describes human nucleic acid-associated proteins
CC   designated NAAP-1 to NAAP-57. The human NAAPs have cytostatic,
CC   antiarteriosclerotic, anti-HIV, antiallergic, cerebroprotective,
CC   antiparkinsonian, anticonvulsant, nootropic, neuroprotective,
CC   antiinflammatory, ophthalmological, thyromimetic, antiarthritic,
CC   hepatotropic, antibacterial, virucide, protozoacide, antiparasitic and
CC   fungicide activities, and can be used in gene therapy. The NAAP protein
CC   and polynucleotide sequences can be used in diagnosing, treating and
CC   preventing diseases or conditions associated with the decreased
CC   expression or overexpression of NAAP, such as cell proliferative diseases
CC   (e.g. cancer, atherosclerosis, hepatitis), neurological disorders
CC   (Parkinson's disease, Alzheimer's disease, stroke, epilepsy),
CC   developmental disorders (renal tubular acidosis, anaemia, glaucoma,
CC   hypothyroidism), autoimmune/inflammatory disorders (AIDS, allergies,
CC   atopic dermatitis, arthritis) and infections (e.g. bacterial, viral,
CC   parasitic, protozoal, fungal). The present sequence represents human NAAP
CC   -41, from the present invention.
XX
SQ   Sequence 217 AA;

Query Match   100.0%;   Score 1134; DB 7;   Length 217;
Best Local Similarity 100.0%;   Pred. No. 46-80;   Indels   0;   Gaps   0;
Matches 212; Conservative   0;   Mismatches   0;

```
Qy    1  SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6  SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA   65

Qy   61  AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66  AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP  125

Qy  121  WAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126  WAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR  185

Qy  181  STILSMSRTWTCRRWAVAPCRABKLMCSSSRS  212
         ||||||||||||||||||||||||||||||||
Db  186  STILSMSRTWTCRRWAVAPCRABKLMCSSSRS  217
```

RESULT 3
AEA20112
ID   AEA20112 standard; protein; 217 AA.
XX
AC   AEA20112;
XX
DT   11-AUG-2005 (first entry)
XX
DE   Novel human polypeptide SEQ ID NO 806.
XX
KW   vulnerary; CNS-gen.; gene therapy; diagnostic; forensic; mapping;
KW   DNA purification; protein purification; osteoarthritis; antiarthritic;
KW   osteopathic; musculoskeletal disease; osteoporosis; endocrine disease;
KW   periodontal disease; antiinflammatory; mouth disease; burns; injury;
KW   peripheral neuropathy; Alzheimers disease; neuroprotective; nootropic;
KW   degeneration; parkinsons disease; antiparkinsonian; neurological disease;
KW   cerebrovascular ischemia; cerebroprotective; vasotropic;
KW   cardiovascular disease; autoimmune disease; immunosuppressive;
KW   immune disorder; viral infection; virucide; infection; cancer;
KW   cytostatic; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO2005049806-A2.
XX
PD   02-JUN-2005.
XX
PF   11-MAR-2004; 2004WO-US007412.
XX
PR   14-MAR-2003; 2003US-00389559.
XX

XX
PA (NUVE-) NUVELO INC.
XX
PI Tang TY, Wang J, Wang ZW, Zhang J, Ren F, Zhou P, Ma Y;
PI Ghosh M, Xue A, Asundi V, Zhao Q, Wang D, Goodrich R, Chen R;
PI Wehrman T, Weng G, Boyle B;
XX
DR WPI; 2005-417730/42.
DR N-PSDB; AEA19545.
XX
PT New polynucleotide encoding a polypeptide with biological activity,
PT useful for treating a disease or disorder, e.g. osteoarthritis, burns,
PT CNS and peripheral disease, stroke, autoimmune disorders, viral
PT infection, or cancer.
XX
PS Claim 20; SEQ ID NO 806; 500pp; English.
XX
CC The invention describes a new isolated polynucleotide (I) encoding a
CC polypeptide with biological activity comprising: a nucleotide sequence of
CC SEQ ID NOS: 1-567 (fully defined); a nucleotide sequence that hybridizes
CC to the sequence of (i) under stringent hybridization conditions; or a
CC nucleotide sequence having greater than 99% sequence identity with the
CC sequence of (i). Also described are: a(n) (expression)vector comprising
CC (I); a host cell genetically engineered to comprise (I) operatively,
CC associated with a regulatory sequence that modulates expression of the
CC polynucleotide in the host cell; an isolated polypeptide comprising a
CC sequence of SEQ ID NOS: 568-1134 (fully defined), where the polypeptide
CC is: a polypeptide encoded by (I); or a polypeptide encoded by a
CC polynucleotide hybridizing under stringent conditions with any one of SEQ
CC ID NOS: 1-567; a composition comprising the polypeptide of (3); and a
CC carrier; an antibody directed against the polypeptide of (3); a method
CC for detecting (I) in a sample; a method for detecting the polypeptide of
CC (3) in a sample; a method for identifying a compound that binds to the
CC polypeptide of (3); a method of producing the polypeptide of (3); and a
CC collection of polynucleotides, where the collection comprising of at
CC least one of SEQ ID NOS: 1-567. (I) is a polynucleotide comprising any of
CC the sequences of SEQ ID NOS: 1-567 encoding a polypeptide with biological
CC activity, which comprises any of the amino acid sequence of SEQ ID NOS:
CC 568-1134. All sequences are fully defined in the specification. The
CC sequences and methods are useful in diagnostics, forensic, and gene
CC mapping, in identifying of mutations responsible for genetic disorders or
CC other traits, in assessing biodiversity, and for producing many other
CC types of data and products dependent on DNA and amino acid sequences. The
CC composition and method are useful for treating a disease or disorder,
CC e.g. osteoporosis, osteoarthritis, periodontal disease, burns, CNS and
CC peripheral disease, Alzheimer's disease, Parkinson's disease, stroke,
CC autoimmune disorders, viral infection, or cancer. This is the amino acid
CC sequence of a novel polypeptide of the invention.
XX
SQ Sequence 217 AA;

Query Match   100.0%; Score 1134; DB 9; Length 217;
Best Local Similarity   100.0%; Pred. No. 4e-80;
Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 65

Qy   61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 AGIASSAVEAVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP 125

Qy   121 WAAAASATPMLSSKASLCIPTRGPPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 WAAAASATPMLSSKASLCIPTERPPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR 185

Qy   181 STILSMSRTWTCRRWAVAPCRABKLMCSSSRS 212
              ||||||||||||||||||||||||||||||||
Db   186 STILSMSRTWTCRRWAVAPCRABKLMCSSSRS 217

RESULT 4

ABU60984
ID ABU60984 standard; protein; 213 AA.
XX
AC ABU60984;
XX
DT 08-MAY-2003 (first entry)
XX
DE Lung specific protein (LSP) #87.
XX
KW Human; gene therapy; vaccine; lung specific antigen; cancer diagnosis;
KW cancer monitoring; cancer staging; cancer imaging; lung cancer;
KW non-cancerous diseases of the lung; transgenic animal.
XX
OS Homo sapiens.
XX
PN WO20026633-A2.
XX
PD 06-SEP-2002.
XX
PF 21-NOV-2001; 2001WO-US043612.
XX
PR 22-NOV-2000; 2000US-0252500P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Macina RA, Recipon H, Chen S, Sun Y, Liu C;
XX
DR WPI; 2002-713376/77.
XX
PT New isolated human nucleic acid molecule and polypeptide, useful for
PT identifying, diagnosing, monitoring, staging, imaging and treating lung
PT cancer and non-cancerous diseases of the lung.
XX
PS Claim 11; Page 368; 389pp; English.
XX
CC The invention describes an isolated human nucleic acid (I) encoding any
CC of 120 10-1533 residue amino acid sequences (S1), given in the
CC specification, comprising any of 164 179-12421 base pair sequences (S2),
CC given in the specification. The methods and compositions of the present
CC invention are useful for identifying, diagnosing, monitoring, staging,
CC imaging and treating lung cancer and non-cancerous diseases of the lung.
CC They are also used for identifying lung tissue, monitoring and
CC identifying and/or designing antagonists of the polypeptide of the
CC invention, gene therapy, Production of transgenic animals and production
CC of engineered lung tissue for treatment and research. This is the amino
CC acid sequence of a lung specific nucleic acid
XX
SQ Sequence 213 AA;

Query Match   94.3%; Score 1069.5; DB 5; Length 213;
Best Local Similarity   95.3%; Pred. No. 4e-75;
Matches 203; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

Qy   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60

Qy   61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMA-SGSSQ 119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| :||
Db   61 AGIASSAVEAVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPTGMARRAASQ 120

Qy   120 PWAAAASATPMLSSKASLCIPTRGPPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAP 179
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 PWAAAASATPMLSSKASLCIPTERPPPPQPLMRTPAARSHWPIPHPASTACPAPLPVVLVAP 180

Qy   180 RSTILSMSRTWTCRRWAVAPCRABKLMCSSSRS 212
              |||||||||||||||||||||||||||||||||
Db   181 RSTILSMSRTWTCRRWAVAPCRABKLMCSSSRS 213

RESULT 5
ABU60939
ID ABU60939 standard; protein; 213 AA.

XX
AC   ABU60939;
XX
DT   08-MAY-2003 (first entry)
XX
DE   Lung specific protein (LSP) #42.
XX
KW   Human; gene therapy; vaccine; lung specific antigen; cancer diagnosis;
KW   cancer monitoring; cancer staging; cancer imaging; lung cancer;
KW   non-cancerous diseases of the lung; transgenic animal.
XX
OS   Homo sapiens.
XX
PN   WO200268633-A2.
XX
PD   06-SEP-2002.
XX
PF   21-NOV-2001; 2001WO-US043612.
XX
PR   22-NOV-2000; 2000US-0252500P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Macina RA, Recipon H, Chen S, Sun Y, Liu C;
XX
DR   WPI; 2002-713376/77.
XX
PT   New isolated human nucleic acid molecule and polypeptide, useful for
PT   identifying, diagnosing, monitoring, staging, imaging and treating lung
PT   cancer and non-cancerous diseases of the lung.
XX
PS   Claim 11; Page 335-336; 389pp; English.
XX
CC   The invention describes an isolated human nucleic acid (I) encoding any
CC   of 120 10-1533 residue amino acid sequences (S1), given in the
CC   specification, comprising any of 164 179-12421 base pair sequences (S2),
CC   given in the specification. The methods and compositions of the present
CC   invention are useful for identifying, diagnosing, monitoring, staging,
CC   imaging and treating lung cancer and non-cancerous diseases of the lung.
CC   They are also used for identifying lung tissue, monitoring and
CC   identifying and/or designing antagonists of the polypeptide of the
CC   invention, gene therapy, production of transgenic animals and production
CC   of engineered lung tissue for treatment and research. This is the amino
CC   acid sequence of a lung specific nucleic acid
XX
SQ   Sequence 213 AA;

Query Match   94.3%; Score 1069.5; DB 5; Length 213;
Best Local Similarity   95.3%; Pred. No. 4e-75;
Matches 203; Conservative   1; Mismatches   8; Indels   1; Gaps   1;

Qy   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60

Qy   61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMA-SGSSQ 119
               |||||||||.||||||||||||||||||||||||||||||||||||| :|| :||
Db   61 AGIASSAVEAVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPTGGMARRAASQ 120

Qy   120 PWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAP 179
               ||||||||||||||||||||| |||||||||||||||||||||| ||||||||||||
Db   121 PWAAASATPMLSSKASLCIPTERPPPQPLMRTPAARSHWPIPHPASTACPAPLPVVLVAP 180

Qy   180 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 212
               |||||||||||||||||||||||||||||||||
Db   181 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 213

RESULT 6
ABJ05501
ID   ABJ05501 standard; protein; 218 AA.
XX
AC   ABJ05501;

XX
DT   14-NOV-2002 (first entry)
XX
DE   Human breast cancer associated polypeptide SEQ ID NO: 261.
XX
KW   Human; breast specific gene; breast specific protein; breast cancer;
KW   gene therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200264611-A1.
XX
PD   22-AUG-2002.
XX
PF   12-FEB-2002; 2002WO-US004197.
XX
PR   13-FEB-2001; 2001US-0268292P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
PI   Sun Y, Liu C;
XX
DR   WPI; 2002-657582/70.
XX
PT   New breast specific nucleic acids and proteins, useful for identifying,
PT   diagnosing, monitoring, staging, imaging, and treating breast cancer and
PT   non-cancerous disease states in breast tissue, and in gene therapy.
XX
PS   Claim 11; Page 339-340; 367pp; English.
XX
CC   The present invention provides human breast specific coding sequences and
CC   proteins. These can be used in the diagnosis and treatment of breast
CC   cancer and non-cancerous diseases of the breast. The present sequence is
CC   a polypeptide of the invention
XX
SQ   Sequence 218 AA;

Query Match   94.0%; Score 1065.5; DB 5; Length 218;
Best Local Similarity   95.3%; Pred. No. 8.3e-75;
Matches 203; Conservative   1; Mismatches   8; Indels   1; Gaps   1;

Qy   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 65

Qy   61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMA-SGSSQ 119
               |||||||||.||||||||||||||||||||||||||||||||||| :|| :||
Db   66 AGIASSAVERVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVQPPAGGMARRAASQ 125

Qy   120 PWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAP 179
               ||||||||||||||||||||| |||||||||||||||||||||| ||||||||||||
Db   126 PWAAASATPMLSSKASLCIPTERPPPQPLMRTPAARSHWPIPHPASTACPAPLPVVLVAP 185

Qy   180 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 212
               |||||||||||||||||||||||||||||||||
Db   186 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 218

RESULT 7
ADQ67157
ID   ADQ67157 standard; protein; 559 AA.
XX
AC   ADQ67157;
XX
DT   07-OCT-2004 (first entry)
XX
DE   Novel human protein sequence #2130.
XX
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   gene therapy; diagnostic marker; morbid state; osteoporosis;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW   cancer.

```
XX
OS   Homo sapiens.
XX
PN   EP1440981-A2.
XX
PD   28-JUL-2004.
XX
PF   21-JAN-2004; 2004EP-00001196.
XX
PR   21-JAN-2003; 2003JP-00102206.
PR   09-MAY-2003; 2003JP-00131392.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI   Yamamoto J, Isono Y, Nagai K, Irie R;
XX
DR   WPI; 2004-535376/52.
DR   N-PSDB; ADQ64969.
XX
PT   Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 4318; 2449pp; English.
XX
CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotide and protein polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a protein
CC   sequence of the invention.
XX
SQ   Sequence 559 AA;

Query Match             13.6%;  Score 154;  DB 8;  Length 559;
Best Local Similarity   25.7%;  Pred. No. 0.0013;
Matches  69;  Conservative  27;  Mismatches  92;  Indels  80;  Gaps  12;

Qy       2  PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS  58
              |  ||:   ||  ||:    |: ||:|   |||
Db      97  PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVVPAAPGPPSLAPSGASPSAS 152

Qy      59  A-AAGIA--------------------SSAVBPVCGDA-APACLLRTPLRGLL  89
              |:|                          |  |  :|
Db     153  ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 212

Qy      90  KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP----- 144
Db     213  NPAPAQASLLAP---------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 259

Qy     145  ---PQPLMRTPAARSHWPIP--HPCDT------ACPAPLPVVLVAPRST------ 182
                      :   ::|                  ::||
Db     260  VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 319

Qy     183  -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
              :  ||:    |||
Db     320  PLPSPASTQTL---ALAPALAPTLGGSS 344

RESULT 8
AAB95541
ID   AAB95541 standard; protein; 1127 AA.
XX
AC   AAB95541;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human protein sequence SEQ ID NO:18149.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
```

```
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;  99JP-00248036.
PR   27-AUG-1999;  99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 18149; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 1127 AA;

Query Match             13.6%;  Score 154;  DB 4;  Length 1127;
Best Local Similarity   25.7%;  Pred. No. 0.0026;
Matches  69;  Conservative  27;  Mismatches  92;  Indels  80;  Gaps  12;

Qy       2  PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS  58
              |  ||:   ||  ||:    |: ||:|   |||
Db     665  PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVVPAAPGPPSLAPSGASPSAS 720

Qy      59  A-AAGIA--------------------SSAVBPVCGDA-APACLLRTPLRGLL  89
              |:|                          |  |  :|
Db     721  ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 780

Qy      90  KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP----- 144
Db     781  NPAPAQASLLAP---------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 827

Qy     145  ---PQPLMRTPAARSHWPIP--HPCDT------ACPAPLPVVLVAPRST------ 182
                      :   ::|                  ::||
Db     828  VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 887
```

```
Qy      183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
            : :|:|  |  :|:|| ||| ||
Db      888 PLPSPASTQTL---ALAPALAPTLGGSS 912

RESULT 9
ADJ70433
ID   ADJ70433 standard; protein; 1480 AA.
XX
AC   ADJ70433;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human heat mitochondrial protein as a therapeutic target SeqID2239.
XX
KW   mitochondrial; human; screening assay; diabetes mellitus;
KW   Huntington's disease; osteoarthritis;
KW   Leber's hereditary optic neuropathy; LHON;
KW   mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW   myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW   neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW   osteopathic; ophthalmological; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200307768-A2.
XX
PD   23-OCT-2003.
XX
PF   04-APR-2003; 2003WO-US010870.
XX
PR   12-APR-2002; 2002US-0372843P.
PR   17-JUN-2002; 2002US-0389987P.
PR   20-SEP-2002; 2002US-0412418P.
XX
PA   (MITO-) MITOKOR.
PA   (BUCK-) BUCK INST AGE RES.
XX
PI   Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI   Warnock DE;
XX
DR   WPI; 2003-845369/78.
XX
PT   Identifying a mitochondrial target for drug screening assays and for
PT   treating diseases associated with altered mitochondrial function,
PT   comprises detecting a modified polypeptide in a sample and correlating
PT   with the disease.
XX
PS   Claim 1; SEQ ID NO 2239; 180pp; English.
XX
CC   This invention relates to novel mitochondrial targets that can be used
CC   for therapeutic intervention in treating a disease associated with
CC   altered mitochondrial function. Specifically, it refers to a method for
CC   identifying proteins of the human heart mitochondrial proteome that are
CC   useful for drug screening assays, as well as therapeutic targets. The
CC   present invention describes a method for identifying such proteins that
CC   can be used in the treatment of various diseases associated with altered
CC   mitochondrial function including diabetes mellitus, Huntington's disease,
CC   osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC   encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC   ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC   compositions have neuroprotective, nootropic, antidiabetic,
CC   anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC   cyostatic activities. This polypeptide sequence is a human heart
CC   mitochondrial protein of the invention.
XX
SQ   Sequence 1480 AA;

Query Match             13.6%;  Score 154; DB 7;  Length 1480;
Best Local Similarity   25.7%;  Pred. No. 0.0035;
Matches   69;  Conservative  27;  Mismatches  92;  Indels  80;  Gaps  12;
```

```
Qy        2 PHQAAAPVD---QTPRTLATMGQRALPSSLALLSRPLSPPAACSGDPGCGSGAGLPSAS 58
            ||:|:      :|  |||           |: : |||||     || ||| :||  |||
Db     1018 PAPASAPLTIPISAPLTVSASG---PALLTSVTPPLAPVVPAAPGPPSLAPSGASPSAS 1073

Qy       59 A-AAGIA-------------------SSAVEPVCGDA-APACLLRTPLRGLL 89
            | :|:                     :||:||  :||:|
Db     1074 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPPPSAP 1133

Qy       90 KPTGPRSTMECPPALIVHPPAGGMASGSQPWAAAASATPMLSSKASLCIPTRGPP----- 144
               :|||:      :||:        :||   ||   |   ||||| |     | ||
Db     1134 NPAPAQASLLAP----------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1180

Qy      145 --PQPLMRTPAARSHWPIP--HPCDT-------ACPAPLPVVLVAPRST------------ 182
              : |    |      ||:|  ||   |       ||:| :| | ||||
Db     1181 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1240

Qy      183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
             : :|:|  |  :|:|| ||| ||
Db     1241 PLPSPASTQTL---ALAPALAPTLGGSS 1265

RESULT 10
ADR09758
ID   ADR09758 standard; protein; 2427 AA.
XX
AC   ADR09758;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Human protein useful for treating neurological disease Seq 3264.
XX
KW   human; oligo-capping method; diagnostic marker; gene therapy;
KW   osteoporosis; neurological disease; Alzheimer's disease;
KW   Parkinson's disease; dementia; short memory; cancer;
KW   sense or motor function; emotional reaction; fear response; panic;
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   tranquilliser.
XX
OS   Homo sapiens.
XX
PN   EP1447413-A2.
XX
PD   18-AUG-2004.
XX
PF   12-FEB-2004; 2004EP-00003145.
XX
PR   14-FEB-2003; 2003JP-00102207.
PR   09-MAY-2003; 2003JP-00131452.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI   Wakamatsu A, Ishii S, Nagai K, Irie R;
XX
DR   WPI; 2004-583265/57.
DR   N-PSDB; ADR07802.
XX
PT   New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 3264; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
CC   cDNA clones obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes an immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
```

```
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polypeptide is a protein
CC   encoded by a full length human cDNA sequence of the invention. NOTE: This
CC   sequence is not given in the sequence listing of the specification but
CC   can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC   office.
XX
SQ   Sequence 2427 AA;

Query Match             13.6%;  Score 154;  DB 8;  Length 2427;
Best Local Similarity   25.7%;  Pred. No. 0.0058;
Matches   69;  Conservative   27;  Mismatches   92;  Indels   80;  Gaps   12;

Qy      2  PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS   58
           ||:||:         ||:  ||  ::|:  ||| ||  ||   ||||||||||||||
Db   1448  PAPASAPLTIPISAPLTVSASG---PALLTSVTPPLAPVVPAAPGPPSLAPSGGASPSAS 1503

Qy     59  A-AAGIA----------------------------SSAVBPVCGDA-APACLLRTPLRGLL   89
           | ||:|                          :|:|| || |||:||||||||
Db   1504  ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1563

Qy     90  KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGPP-----  144
           ||| || |      |||  |||| ||||||  ||  || || ||  : :: |
Db   1564  NPAPAQASLLAP------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1610

Qy    145  ---PQPLMRTPAARSHWPIP--HPCDT---------ACPAPLPVVLVAPRST---------  182
              |   | ||||||| |||  ||||||         |||||||||||||  ||
Db   1611  VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1670

Qy    183  -ILSMSRTWTCRRWAVAPCRAEKLMCSS  209
            ||||||: || | | | ||:||||||
Db   1671  PLPSPASTQTL--ALAPALAPTLGGSS 1695
```

RESULT 11
```
ABR41356
ID   ABR41356 standard; protein; 2735 AA.
XX
AC   ABR41356;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human DITHP transcription factor.
XX
KW   Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW   cancer; cell proliferative disorder; autoimmune disorder;
KW   inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW   neurological disorder; gastrointestinal disorder; transport disorder;
KW   connective tissue disorder; drug screening; proteome analysis;
KW   gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW   disease model; toxicological testing; transcript imaging;
KW   transcription factor.
XX
OS   Homo sapiens.
XX
PN   WO200297031-A2.
XX
PD   05-DEC-2002.
XX
PF   27-MAR-2002; 2002WO-US010056.
XX
PR   28-MAR-2001; 2001US-0279619P.
PR   29-MAR-2001; 2001US-0280067P.
PR   29-MAR-2001; 2001US-0280068P.
PR   16-MAY-2001; 2001US-0291280P.
PR   17-MAY-2001; 2001US-0291829P.
PR   17-MAY-2001; 2001US-0291849P.
PR   19-JUN-2001; 2001US-0299428P.
PR   20-JUN-2001; 2001US-0299776P.
PR   20-JUN-2001; 2001US-0300001P.
XX
```

PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI   Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PB, Amshey SR;
PI   Daughtery, SC, Dam TC, Liu TP, Nguyen DA, Kleefeld Y, Gerstin EH;
PI   Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI   Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
```
DR   WPI; 2003-129518/12.
DR   N-PSDB; ACC46296.
XX
PT   Novel human diagnostic and therapeutic polypeptide useful for identifying
PT   test compound which specifically binds to a polypeptide encoded by human
PT   diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
PS   Claim 27; SEQ ID NO 891; 591pp; English.
XX
CC   The invention relates to novel human diagnostic and therapeutic
CC   polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC   proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC   polynucleotide sequences at least 90% identical to the dithp cDNA
CC   sequences of the invention; recombinant vectors, host cells and
CC   transgenic organisms comprising a dithp nucleic acid sequence; the
CC   recombinant production of DITHP proteins; antibodies specific for DITHP
CC   proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC   detecting dithp nucleotide and protein sequences; methods of screening
CC   for compounds which specifically bind a DITHP protein; and methods of
CC   assessing the toxicity of test compounds using a dithp hybridisation
CC   probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC   diagnosis of a wide variety of conditions including cancer and other cell
CC   proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC   viral, fungal or parasitic infections; hormonal disorders; metabolic
CC   disorders; neurological disorders; gastrointestinal disorders; transport
CC   disorders; and connective tissue disorders. They may also be used to
CC   screen for modulators of protein activity or gene expression. DITHP
CC   proteins can additionally be used in analysis of the proteome of a tissue
CC   or cell type and to induce antibodies. The dithp nucleic acids are
CC   additionally useful in somatic or germline gene therapy of the disorders
CC   mentioned above, as a source of antisense sequences, as a source of
CC   probes and primers, in genotyping and identification of individuals, in
CC   the generation of transgenic animal models of human disease or knock in
CC   humanised animals, in toxicological testing, and in transcript imaging.
CC   The present sequence represents a DITHP protein which has transcription
CC   factor activity. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2735 AA;

Query Match             13.6%;  Score 154;  DB 6;  Length 2735;
Best Local Similarity   25.7%;  Pred. No. 0.0066;
Matches   69;  Conservative   27;  Mismatches   92;  Indels   80;  Gaps   12;

Qy      2  PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS   58
           ||:||:         ||:  ||  ::|:  ||| ||  ||   ||||||||||||||
Db    953  PAPASAPLTIPISAPLTVSASG---PALLTSVTPPLAPVVPAAPGPPSLAPSGGASPSAS 1008

Qy     59  A-AAGIA----------------------------SSAVBPVCGDA-APACLLRTPLRGLL   89
           | ||:|                          :|:|| || |||:||||||||
Db   1009  ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1068

Qy     90  KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGPP-----  144
           ||| || |      |||  |||| ||||||  ||  || || ||  : :: |
Db   1069  NPAPAQASLLAP------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1115

Qy    145  ---PQPLMRTPAARSHWPIP--HPCDT---------ACPAPLPVVLVAPRST---------  182
              |   | ||||||| |||  ||||||         |||||||||||||  ||
Db   1116  VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1175

Qy    183  -ILSMSRTWTCRRWAVAPCRAEKLMCSS  209
            ||||||: || | | | ||:||||||
Db   1176  PLPSPASTQTL---ALAPALAPTLGGSS 1200
```

RESULT 12
AAB41231
ID   AAB41231 standard; protein; 2971 AA.
XX
AC   AAB41231;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human ORFX ORF995 polypeptide sequence SEQ ID NO:1990.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; antifungal; antirheumatic; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US008621.
XX
PR   31-MAR-1999;   99US-0127607P.
PR   02-APR-1999;   99US-0127636P.
PR   05-APR-1999;   99US-0127728P.
PR   30-MAR-2000;   2000US-00540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; AAC75440.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease.
XX
PS   Claim 11; Page 1510-1517; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive

SQ   Sequence 2971 AA;

Query Match   13.6%; Score 154; DB 3; Length 2971;
Best Local Similarity   25.7%; Pred. No. 0.0072;
Matches   69; Conservative   27; Mismatches   92; Indels   80; Gaps   12;

Qy   2 PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS   58
                         :|:| |:||:|:|||:||| ||||:|| |||||
Db   1189 PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVPAAPGPPSLAPSGASPSAS 1244

Qy   59 A-AAGIA-----------------------------SSAVEPVCGDA-APACLLRTPLRGLL   89
               | |:| :|:| |:|:|: |:|:|
Db   1245 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1304

Qy   90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP----- 144
                   :::| |:|||:|||||:|||:|:||:|:||:|
Db   1305 NPAPAQASLLAP-----------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1351

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT------ACPAPLPVVLVAPRST------------ 182
                  ||||:|:|:||:||:|: ||:| :|:|| |:||:|||
Db   1352 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1411

Qy   183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
              :|::| :|||::| |:|:|: |:|||
Db   1412 PLPSPASTQTL---ALAPALAPTLGGSS 1436

RESULT 13
ADC35084
ID   ADC35084 standard; protein; 2971 AA.
XX
AC   ADC35084;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Human breast cancer antigen seq id 50.
XX
KW   breast cancer; breast cancer diagnosis; breast cancer antigen.
XX
OS   Homo sapiens.
XX
PN   US2003108888-A1.
XX
PD   12-JUN-2003.
XX
PF   15-MAY-2002; 2002US-00146473.
XX
PR   15-MAY-2001; 2001US-0291150P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Scanlan MJ, Gout I, Stockert E, Old LJ, Gure A, Chen Y;
XX
DR   WPI; 2003-829397/77.
DR   N-PSDB; ADC35126.
XX
PT   Diagnosing breast cancer in subject by obtaining biological sample from
PT   subject, contacting sample with breast cancer-associated polypeptides,
PT   determining specific binding between polypeptides and agents in sample.
XX
PS   Example 2; SEQ ID NO 50; 173pp; English.
XX
CC   The invention describes a method of diagnosing breast cancer in subject
CC   comprising contacting biological sample from subject with at least two
CC   different breast cancer-associated polypeptides (I) encoded by nucleic
CC   acid molecules (II) comprising sequence chosen from 42 fully defined
CC   sequences as given in specification, determining specific binding between
CC   (I) and agents in sample, where presence of the binding is diagnostic for
CC   breast cancer. The method is useful for diagnosing breast cancer in a
CC   subject. The sample is blood, lymph node fluid or breast discharge fluid.
CC   This is the amino acid sequence of a breast cancer antigen.
XX
SQ   Sequence 2971 AA;

```
Query Match        13.4%;    Score 152; DB 7;   Length 2971;
Best Local Similarity    25.7%;    Pred. No. 0.01;
Matches    69;  Conservative    27;  Mismatches    92;  Indels    80;  Gaps    12;

Qy     2 PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS    58
           | :||:      ||::    :   ||:||    :::|| |||
Db  1189 PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVVPAAPGPPSLQPSGASPSAS 1244

Qy    59 A-AAGIA----------------SSAVBPVCGDA-APACLLRTPLRGLL    89
          |:|:                  |:
Db  1245 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1304

Qy    90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGPP----- 144
          ::::|                 |
Db  1305 NPAPAQASLLAP-------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1351

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT------ACPAPLPVLVAPRST----------- 182
             :|:         :||:       ::|||
Db  1352 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1411

Qy   183 -ILSMSRTWTCRRWAVAPCRABKLMCSS 209
           :|:||::          |
Db  1412 PLPSPASTQTL--ALAPALAPTLGGSS 1436

RESULT 14
AAB50363
ID  AAB50363 standard; protein; 2972 AA.
XX
AC  AAB50363;
XX
DT  12-MAR-2001  (first entry)
XX
DE  Human SRCAP.
XX
KW  Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;
KW  cAMP regulatory element; CREB binding protein; CBP; ATPase;
KW  transcription activation; DEAD box RNA dependent helicase;
KW  adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.
XX
OS  Homo sapiens.
XX
PN  WO200073467-A1.
XX
PD  07-DEC-2000.
XX
PF  25-MAY-2000; 2000WO-US014719.
XX
PR  27-MAY-1999;   99US-0136620P.
PR  25-MAY-2000; 2000US-00579181.
XX
PA  (UYSL-) UNIV SAINT LOUIS.
XX
PI  Chrivia J, Yaciuk P;
XX
DR  WPI; 2001-061545/07.
DR  N-PSDB; AAC89860.
XX
PT  Snf2 related cAMP regulatory element (CREB) binding protein (CBP)
PT  activator protein, capable of co-activating CREB binding protein, useful
PT  for modulating transcription and for affecting viral infection.
XX
PS  Claim 1; Page 86-94; 103pp; English.
XX

CC  The present sequence is an Snf2 related CREB (cAMP regulatory element)
CC  binding protein (CBP) activator protein (SRCAP) polypeptide. It has
CC  ATPase activity and is capable of activating transcription. SRCAP
CC  polypeptides are useful for activating transcription in a cell, for
CC  enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated
CC  activation of transcription in a cell, for treating a patient having a
CC  disease involving a function such as insufficient transcription of a
CC  gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent
CC  helicase, adenoviral DBP protein, beta-actin or a nuclear receptor
CC  affected by SRCAP protein. Compounds that modulate SRCAP function, such
CC  as antibodies, antisense molecules, polynucleotides or ribozymes, are
CC  useful for treating diseases mediated by SRCAP-activated transcription,
CC  for example infection by adenovirus, hepatitis C virus, human
CC  immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or
CC  hepatitis B virus
XX
SQ  Sequence 2972 AA;

Query Match        13.4%;    Score 152;  DB 4;   Length 2972;
Best Local Similarity    25.7%;   Pred. No. 0.01;
Matches    69;  Conservative    27;  Mismatches   92;  Indels   80;  Gaps   12;

Qy     2 PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS    58
           | :||:      ||::    :   ||:||    :::|| |||
Db  1189 PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVVPAAPGPPSLQPSGASPSAS 1244

Qy    59 A-AAGIA----------------SSAVBPVCGDA-APACLLRTPLRGLL    89
          |:|:                  |:
Db  1245 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1304

Qy    90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGPP----- 144
          |  ::::|                |
Db  1305 NPAPAQASLLAP-------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1351

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT------ACPAPLPVLVAPRST----------- 182
             :|:         :||:       ::|||
Db  1352 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1411

Qy   183 -ILSMSRTWTCRRWAVAPCRABKLMCSS 209
           :|:||::          |
Db  1412 PLPSPASTQTL--ALAPALAPTLGGSS 1436

RESULT 15
AAB50362
ID  AAB50362 standard; protein; 3118 AA.
XX
AC  AAB50362;
XX
DT  12-MAR-2001  (first entry)
XX
DE  Human SRCAP.
XX
KW  Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;
KW  cAMP regulatory element; CREB binding protein; CBP; ATPase;
KW  transcription activation; DEAD box RNA dependent helicase;
KW  adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.
XX
OS  Homo sapiens.
XX
PN  WO200073467-A1.
XX
PD  07-DEC-2000.
XX
PF  25-MAY-2000; 2000WO-US014719.
XX
PR  27-MAY-1999;   99US-0136620P.
PR  25-MAY-2000; 2000US-00579181.
XX
PA  (UYSL-) UNIV SAINT LOUIS.
XX
PI  Chrivia J, Yaciuk P;
XX
DR  WPI; 2001-061545/07.
DR  N-PSDB; AAC89859.
XX
PT  Snf2 related cAMP regulatory element (CREB) binding protein (CBP)
PT  activator protein, capable of co-activating CREB binding protein, useful
PT  for modulating transcription and for affecting viral infection.
XX
PS  Claim 6; Page 77-86; 103pp; English.
```

```
XX
CC    The present sequence is an Snf2 related CREB (cAMP regulatory element)
CC    binding protein (CBP) activator protein (SRCAP) polypeptide. It has
CC    ATPase activity and is capable of activating transcription. SRCAP
CC    polypeptides are useful for activating transcription in a cell, for
CC    enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated
CC    activation of transcription in a cell, for treating a patient having a
CC    disease involving a function such as insufficient transcription of a
CC    gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent
CC    helicase, adenoviral DBP protein, beta-actin or a nuclear receptor
CC    affected by SRCAP protein. Compounds that modulate SRCAP function, such
CC    as antibodies, antisense molecules, polynucleotides or ribozymes, are
CC    useful for treating diseases mediated by SRCAP-activated transcription,
CC    for example, infection by adenovirus, hepatitis C virus, human
CC    immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or
CC    hepatitis B virus
XX
SQ    Sequence 3118 AA;

Query Match             13.4%;  Score 152;  DB 4;  Length 3118;
Best Local Similarity   25.7%;  Pred. No. 0.011;
Matches   69;  Conservative 27;  Mismatches  92;  Indels  80;  Gaps  12;

Qy         2 PHQAAAPVD---QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS   58
             |  :|:||    ||:  |  | ::| ||| |  |||| ||  |:| ||:|  ||||
Db      1335 PAPASAPLTIPISAPLTVSASG---PALLTSVTPPLAPVVPAAPGPPSLQPSGASPSAS 1390

Qy        59 A-AAGIA---------------------------SSAVEPVCGDA-APACLRTPLRGLL   89
             |  | |:|                            :|   :| : |  |  | |
Db      1391 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPPPSAP 1450

Qy        90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP-----  144
             |   :::|  |  ||||     ||||   | || :|: ::  |: :  || :| ||
Db      1451 NPAPAQASLLAP----------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1497

Qy       145 ---PQPLMRTPAARSHWPIP--HPCDT------ACPAPLPVVLVAPRST----------  182
                 | | |  |::||      :|:|         |:|| :||:  ||:
Db      1498 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1557

Qy       183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS  209
              :|  |||| |  |:|| || ::| |||
Db      1558 PLPSPASTQTL---ALAPALAPTLGGSS 1582
```

Search completed: March 16, 2006, 15:52:29
Job time : 191 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: March 16, 2006, 15:52:45 ; Search time 40 Seconds
(without alignments)
509.949 Million cell updates/sec

Title: US-09-989-890-238
Perfect score: 1134
Sequence: 1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRABKLMCSSSRS 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	154.5	13.6	1952	2	T48814	hypothetical prote
2	136.5	12.0	635	2	F75477	hypothetical prote
3	135.5	11.9	805	2	T49385	hypothetical prote
4	135	11.9	3164	1	WMBEH6	UL36 protein - hum
5	134.5	11.9	1721	1	I38902	retinoblastoma bin
6	133	11.7	873	2	A47283	calphotin - fruit
7	132	11.6	574	2	T43556	Wiskott-Aldrich sy
8	132	11.6	574	2	T38819	wiskott-aldrich sy
9	131	11.6	865	2	A47282	calcium-binding pr
10	129.5	11.4	1188	2	S49915	extensin-like prot
11	129.5	11.4	1199	2	A40670	nuclear envelope p
12	129.5	11.4	3149	1	QQBE8	BPLF1 protein - hu
13	129	11.4	580	2	T43481	probable mucin DKF
14	129	11.4	1006	2	G86292	hypothetical prote
15	128.5	11.3	507	2	T44768	antifreeze glycope
16	126	11.1	559	2	I49444	SH3 binding protei
17	125.5	11.1	576	2	T36729	probable serine/th
18	125.5	11.1	801	2	T29018	hypothetical prote
19	123.5	10.9	1047	2	A55617	masquerade precurs
20	123	10.8	1706	2	I84499	zinc finger protei
21	121.5	10.7	186	2	C41132	collagen-related p
22	120	10.6	349	2	T05857	hypothetical prote
23	119	10.5	1611	2	T38236	hypothetical prote
24	118	10.4	222	2	H96711	hypothetical prote
25	117.5	10.4	383	2	T46707	proteophosphoglyca
26	117	10.3	839	2	T04859	extensin homolog F
27	117	10.3	1299	2	T47182	hypothetical prote
28	117	10.3	2187	2	T30826	nascent polypeptid
29	116.5	10.3	817	2	S51342	verprolin - yeast
30	116.5	10.3	1151	2	T18535	high molecular mas
31	116	10.2	924	2	S27923	gene LF3 protein -
32	115.5	10.2	744	2	T35192	probable ABC trans
33	115.5	10.2	839	2	P75518	hypothetical prote
34	115.5	10.2	981	1	FOMVGM	gag-abl polyprotei
35	115.5	10.2	1008	2	T04462	hypothetical prote
36	115.5	10.2	1069	2	D85383	hypothetical prote
37	115.5	10.2	1123	2	A39962	kinase-related tra
38	115	10.1	1613	2	S39059	protein BRG1 - hum
39	115	10.1	1647	2	S45252	SNF2beta protein -
40	114.5	10.1	1273	2	S58782	SEC31 protein - ye
41	114	10.1	742	2	F84643	hypothetical prote
42	114	10.1	775	1	EDBE11	immediate-early pr
43	114	10.1	5262	2	T03454	ALR protein - huma
44	113.5	10.0	377	2	A48018	mucin 7 precursor,
45	113	10.0	476	2	T27051	hypothetical prote

ALIGNMENTS

RESULT 1
T48814
hypothetical protein 15B6.220 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C;Accession: T48814
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24541
A;Accession: T48814
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1952 <SCH>
A;Cross-references: UNIPROT:Q9P6T1; UNIPARC:UPI000017B444; EMBL:AL353822; GSPDB:GN00112
A;Experimental source: cosmid contig 15B6; strain 74
C;Genetics:
A;Gene: NCSP:15B6.220
A;Map position: 2
A;Introns: 281/3

Query Match   13.6%;   Score 154.5; DB 2; Length 1952;
Best Local Similarity   30.7%;   Pred. No. 0.01;
Matches   62; Conservative   20; Mismatches   77; Indels   43; Gaps   9;

Qy   1 SPHQAAAPVDQT-PRTLATMGQRALP---SSLALLSRPLSPPPAACSGDPGCGSGAGLP 55
            |||:||:|||| ::|:: |||: |::|:|:| :||:|
Db   1761 TPHSASRPVPQSVPQSVPQATPQAVPRPSTSALTPAQPGPVSPAV------SGGSVP 1812

Qy   56 SASAAAGIASSAVB--PV-CGDAAPACLLR---TPLRGLLKP-----TGPRSTMBCP-- 101
            |||:|:|:|||:| :: ||:||| : |:: ||::::|:   :||:|||:|
Db   1813 APSAAQSVAPAPVSSTPVPAATVAPASTVAAAPTPTRVTAAPAALSAATNPAPVPSQPQH 1872

Qy   102 ------PALIVHPPAGGMASGSSQPWAAAGATPMLSSKASLCIPTRGPPPQPLMRTPAAR 155
                  |:|: ::|| :||:|:   :|::|::|::|:|:|
Db   1873 QITGQAPAQQQRPPA----------QAQAPATPTITSAAPPRPPTLAPPPPPPPTED 1922

Qy   156 SHWPIPHPCDTACPAPLPVVLV 177
            |:::|||:||:|| |:|:||:|
Db   1923 PPPPPPPPAEAPPPPPTPLM 1944

RESULT 2
F75477
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: F75477
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

RESULT 4
WMBEH6
UL36 protein - human herpesvirus 1 (strain 17)
C;Species: human herpesvirus 1
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 09-Jul-2004
C;Accession: I30085
R;McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perr...
J. Gen. Virol. 69, 1531-1574, 1988
A;Title: The complete DNA sequence of the long unique region in the genome of herpes sim...
A;Reference number: A30083; MUID:88274327; PMID:2839594
A;Accession: I30085
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3164 <MCG>
A;Cross-references: UNIPROT:P10220; UNIPARC:UPI0000013B9B; GB:X14112; NID:g1944536; PIDN...
C;Genetics:
A;Gene: UL36
C;Superfamily: varicella-zoster virus gene 22 protein

Query Match   11.9%; Score 135; DB 1; Length 3164;
Best Local Similarity   25.7%; Pred. No. 0.29;
Matches 68; Conservative 17; Mismatches 86; Indels 94; Gaps 12;

Qy   1 SPHQAAAPVDQT--------------------------PRTLA----TMGQR--ALP 25
Db 2651 SPQSSPAPPDATAPRPPASSRASAAASSSGRARRHRRARSLARATQASATTQGWRPPALP 2710

Qy   26 SSLALL--SRPLSPPPAACSGDPGCGSGAGLPSASAAAGIASSA--VEPVCGDAAPACL 80
Db 2711 DTVAPVTDFARPPAPPKPPEPAPHALVSGVPLPLGPQAAGQASPALPIDPVPPPVATGTV 2770

Qy   81 L------RTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSS----QPWAAASAT---- 127
Db 2771 LPGGENRRPPLTSGPAPTPPRVPVGGPQRRLTRPAVASLSESRESLPSPWDPADPTAPVL 2830

Qy 128 ------PMLSSKAS-----------LCIPTRGPPPQ----------PLMRTPAARSHW 158
Db 2831 GRNPAEPTSSSPAGPSPPPPAVQPVAPPFTSGPPPTYLTLEGGVAPGGPVSRRPTTRQ-- 2888

Qy 159 PIPHPCDTACP-------------------APLP 173
Db 2889 PVATPTTSARPRGHLTVSRLSAPQP 2913

RESULT 5
I38902
retinoblastoma binding protein RIZ - human
N;Alternate names: retinoblastoma interacting zinc-finger (RIZ) protein; transcription f...
C;Species: Homo sapiens (man)
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 31-Dec-2004
C;Accession: I38902; A59221; PC4095
R;Buyse, I.M.; Shao, G.; Huang, S.
Proc. Natl. Acad. Sci. U.S.A. 92, 4467-4471, 1995
A;Title: The retinoblastoma protein binds to RIZ, a zinc-finger protein that shares an e...
A;Reference number: I38902; MUID:95273384; PMID:7538672
A;Accession: I38902
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 3-1721 <BUY1>
A;Cross-references: UNIPROT:Q13029; UNIPARC:UPI000013217C; EMBL:U17838;
A;Accession: A59221
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1721 <BUY2>
A;Cross-references: UNIPARC:UPI00001740B0; EMBL:U17838; NID:g1669774; PIDN:AAC50820.1; P...
A;Note: the sequence is revised in GenBank entry HSU17838, release 114.0, PIDN:AAC50820.
R;Shapiro, V.S.; Lee, P.; Winoto, A.
Gene 163, 329-330, 1995
A;Title: Identification and cloning of the G3B cDNA encoding a 3' segment of a protein b...
A;Reference number: PC4095; MUID:96011660; PMID:7590293
A;Accession: PC4095
A;Molecule type: mRNA
A;Residues: 161-165,'T',167-197,'S',199-201,'G',203-277,'G',296-308,'SC...
A;Residues: 161-165,'T',167-197,'S',199-201,'G',203-277,'VGGGGGVVVVSWKARGE',296-308,'SC...

A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75477
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-635 <WHI>
A;Cross-references: UNIPROT:Q9RW96; UNIPARC:UPI00000C1804; GB:AE001932; NID...
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0773
A;Map position: 1

Query Match   12.0%; Score 136.5; DB 2; Length 635;
Best Local Similarity   26.6%; Pred. No. 0.055;
Matches 61; Conservative 29; Mismatches 78; Indels 61; Gaps 11;

Qy   1 SPHQAAAPVDQTPRTLATMGQRALPSSLA-----LLSRPLSP-PPAACS-------GDP 46
Db 264 SPTEVRSPVTSAPVSGSSVNGDSVGSGMATAPTVVVSTPAQPTTTPSASSPVARNTAGQS 323

Qy 47 GCG--SGAGLP----------------------SASAAAGIASSAVEPVCGDAAPACLLR 82
Db 324 GAGQNTGASRPASGTPSRASAPNVVVRPPARVSGTSAAGATPMLSSKASL----- A 378

Qy 83 TPLRGLLKPTGPR-STMECPPALIVHPPAGGM--ASGSSQPWAAASATPMLSSKASL-- 136
Db 379 TPTRN-QSASGTRPASGQTPPRVIAVAPAGGARTPAAPGNRPATAAPTRPSPSSTAPSKT 437

Qy 137 --CIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR 180
Db 438 VVVRPPVPVSGQQPPRVVTSPAAGV------PAGNAAPAPTPPPVVAAR 480

RESULT 3
T49385
hypothetical protein B1D1.390 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T49385
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,...
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49385
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-805 <SCH>
A;Cross-references: UNIPROT:Q9P682; UNIPARC:UPI000017B46C; EMBL:AL355927; GSPDB:GN00116;...
A;Experimental source: BAC clone B1D1; strain OR74A
C;Genetics:
A;Gene: NCSP:B1D1.390
A;Map position: 6
A;Introns: 54/3; 212/3

Query Match   11.9%; Score 135.5; DB 2; Length 805;
Best Local Similarity   27.6%; Pred. No. 0.079;
Matches 54; Conservative 24; Mismatches 85; Indels 33; Gaps 6;

Qy   5 AAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAAAGIA 64
Db 333 ARAPLQYPVPYTATPTPPVIPGA------VPPPQVPCGRP-------PSALAVAGPS 377

Qy 65 SSA-----VEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPP-ALIVHPPAGGMASGSS 118
Db 378 AGAGGALSIDSLFGQGALAALAGATRKSATPTPSQTGPTPQPPIATPIPPPAVPPVAGAP 437

Qy 119 QPWAAASA------TPMLSSKASLCIPTRGPPPPQPLMRTPAARSHWPIPHPCDTACPAPL 172
Db 438 TPVPAAAGSLAALPPVIAAVLRSQTPTVGPAAPPLEAPKPAASAP------PTVPAAA 491

Qy 173 PVVLVAPRSTILSMSR 188
Db 492 PPAPVSNPSALLAMLR 507

```
Qy 151 TP--AARSHWPIPHPCDTACPAPLPVV------LVAPRSTILSMSRTWTCRRWAVAPCR 201
           ||   ::: : ::||                |:
Db 168 TPPVVPANTTVPVAAPV-AAVPAAVPVAAPVLAPAVAP--------------AVAPVV 210

Qy 202 AB 203
           ||
Db 211 AB 212
```

RESULT 7
T43556
Wiskott-Aldrich syndrome protein homolog - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T43556
R;Zankel, T.C.; Ow, D.W.
submitted to the EMBL Data Library, December 1997
A;Description: A Wiskott-Aldrich Syndrome protein homolog in Schizosaccharomyces pombe,
A;Reference number: Z22575
A;Accession: T43556
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-574 <ZAN>
A;Cross-references: UNIPROT:O36027; UNIPARC:UPI000003CA2E; EMBL:AF038575; PIDN:AAB92587
A;Experimental source: strain JS21
C;Genetics:
A;Gene: wsp1
A;Map position: I
A;Introns: 72/3; 519/3; 564/1

Query Match   11.6%;   Score 132; DB 2; Length 574;
Best Local Similarity   28.3%;   Pred. No. 0.099;
Matches   68;   Conservative   20;   Mismatches   82;   Indels   70;   Gaps   13;

```
Qy   1 SPHQAAAPVDQTPRTLATMGQRALP------SSLALLS---RPLSPPPAA--CSGDPGCG 49
         |::                       |:           ::::||
Db 270 SPPRPIAPVSMNPAINST-SKPPLPPPSSRVSAAALAANKKRPPPPPPSRRNRGKPPIG 328

Qy  50 SGAG------LPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPR--STM 98
          |:                                                :|
Db 329 NGSSNSSLPPPPPPRSNAAG-----SIPLPPQGRSAP--------PPPPRSAPSTG 373

Qy  99 ECPPAL-----IVHPPA-----GGMASGSGSQPWAAASAT--------PMLSSKASLCIPT 140
           |:|        ||                   : |||
Db 374 RQPPPLSSSRAVSNPPAPPPAIPGRSAPALPLGNASRTSTPPPVPTPPSLPPSAPPSLPP 433

Qy 141 RGPPPQPLMRTPAA------------------RSHWPIPHPCDTACPAPLPVVLVAPRSTILSMS 187
           |:|||                          :|
Db 434 SAPPSLP-MGAPAAPPLPPSAPIAPPLPAGMPAPAAPPLPPAAPAPPAPAAPVASIA 492
```

RESULT 8
T38819
wiskott-aldrich syndrome protein homolog 1 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T38819
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21813
A;Accession: T38819
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-574 <CON>
A;Cross-references: UNIPROT:O36027; UNIPARC:UPI000006C8BE; EMBL:Z98980; NID:e1060691; P
A;Experimental source: strain 972h-; cosmid c4F10
C;Genetics:
A;Gene: wsp1; SPDB:SPAC4F10.15c
A;Map position: 1
A;Introns: 72/3; 519/3; 564/1

Query Match   11.6%;   Score 132; DB 2; Length 574;
Best Local Similarity   28.3%;   Pred. No. 0.099;

A;Cross-references: UNIPARC:UPI000007377B; GB:U23736; NID:g915214; PIDN:AAA87023.1; PID
C;Genetics:
A;Gene: GDB:RIZ; G3B
A;Cross-references: GDB:636678; OMIM:601196
A;Map position: 1p36.23-1p36.13
C;Superfamily: SH3 homology
C;Keywords: zinc finger
F;364-384/Region: zinc finger CCHH motif
F;394-414/Region: zinc finger CCHH motif
F;487-508/Region: zinc finger CCHH motif
F;752-805/Domain: SH3 homology <SH3>
F;1139-1159/Region: zinc finger CCHH motif
F;1167-1188/Region: zinc finger CCHH motif
F;1196-1217/Region: zinc finger CCHH motif
F;1338-1358/Region: zinc finger CCHH motif
F;1460-1481/Region: zinc finger CCHH motif

Query Match   11.9%;   Score 134.5; DB 1; Length 1721;
Best Local Similarity   28.3%;   Pred. No. 0.18;
Matches   54;   Conservative   21;   Mismatches   71;   Indels   45;   Gaps   9;

```
Qy  19 MGQRALPSSLALLSRPLSPP------PAACSG----DPGCGSGAGLPSASAAAGIASSA 67
          |                       :|                    :
Db 890 MLQKVLLNEYNGIDLPVENPADGTRSPSPCKSLBAQPDDDLGPGSGFPAPTV-----ES 943

Qy  68 VEPVCGDAAPACLLRTP---------LRGLLKPTGPRSTMECPPALIVHPPAGGMASGSGSSQPW 121
                                                  |:
Db 944 TPDVC-PSSPA--LQTPSLSSGQLPLLIPTDPSSPPPCPPVLTVATPPPLLPTVPLPA 1000

Qy 122 AAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVV-----L 176
          :::|:                             |
Db 1001 PSSSASP-----------HPCPSPLSNA-TAQSPLPILSPTVSPSPSFSPIPPVEPLMSA 1046

Qy 177 VAPRSTILSMS 187
          :
Db 1047 ASPGPPTLSSS 1057
```

RESULT 6
A47283
calphotin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: A47283
R;Ballinger, D.G.; Xue, N.; Harshman, K.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 1536-1540, 1993
A;Title: A Drosophila photoreceptor cell-specific protein, calphotin, binds calcium and
A;Reference number: A47283; MUID:93165730; PMID:8434015
A;Contents: photoreceptor cells
A;Accession: A47283
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-873 <BAL>
A;Cross-references: UNIPROT:Q02910; UNIPARC:UPI000016BB46; GB:L05080; NID:g157071; PIDN
A;Note: sequence extracted from NCBI backbone (NCBIN:124958, NCBIP:124959)
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match   11.7%;   Score 133; DB 2; Length 873;
Best Local Similarity   27.3%;   Pred. No. 0.12;
Matches   66;   Conservative   21;   Mismatches   21;   Indels   76;   Gaps   13;

```
Qy   1 SPHQA--AAPVDQTPRTLATMGQRALPSSLALLS----------RPLSPPPAACSGD 45
          |:                                         :||
Db   8 SPVSAPVAAPV--TPSAVAAPVQVVSPAAVAVAPAVVAPAPAAPTAVTPVAPPPTLASVQ 65

Qy  46 PGCGSGAGLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALI 105
                                                        :|
Db  66 P----ATVTVPAPAPIAAASVAPVASVAPPVVAAPTP----PAASPVST---PPVAV 111

Qy 106 VHPPAGGMASGSGSQPWAAASATPMLSSKASLCIPTRGPPP---------QPLMR 150
                                                                :::
Db 112 AQIPVA--VSAPVAPPVAATPTPV--APIPVAAPVIATPPVAASAPTPAAVTPVVSPVIA 167
```

```
Matches    68;  Conservative    20;  Mismatches    82;  Indels    70;  Gaps    13;

Qy    1  SPHQAAAPVDQTPRTLATMGQRALP------SSLALLS---RPLSPPPAA--CSGDPGCG   49
         |||   ||||||||||||||||||          : |||    ||| |||||  ||||||||
Db  270  SPRPIAPVSMNPAINST-SKPPLPPPSSRVSAAALAANKKRPPPPPPSRRNRGKPPIG    328

Qy   50  SGAG------LPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPR---STM     98
         ||||      ::|||||||||||||||||||||||||||||||||||||||||
Db  329  NGSSNSSLPPPPPPPRSNAAG---SIPLPPQGRSAP---------PPPPPRSAPSTG     373

Qy   99  ECPPAL-----IVHPPA-----GGMASGSSQPWAAASAT--------PMLSSKASLCIPT   140
         : |:       |||||      :: ||||||||||||
Db  374  RQPPPLSSSRAVSNPPAPPPAIPGRSAPALPPLGNASRTSTPPVFTPPSLPPSAPPSLPP   433

Qy  141  RGPPPQPLMRTPAA-----------RSHWPIPHCDTACPAPLPVVLVAPRSTILSMS     187
         | |||                    |||
Db  434  SAPPSLP-MGAPAPAPLPPSAPIAPPLPAGMPRAPPLPPAAPAPPPAAPAPVASIA     492

RESULT 9
A47282
calcium-binding protein calphotin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A47282
R;Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993
A;Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.
A;Reference number: A47282; MUID:93165729; PMID:8094559
A;Accession: A47282
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-865 <MAR>
A;Cross-references: UNIPROT:Q02910; UNIPARC:UPI00001282D3; GB:L02111; NID:g157031; PIDN:
A;Experimental source: photoreceptor cells
A;Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBIP:124956)
C;Genetics:
A;Gene: FlyBase:Cpn
A;Cross-references: FlyBase:FBgn0010218
A;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: calcium binding

Query Match    11.6%;  Score    131;  DB    2;  Length    865;
Best Local Similarity    26.6%;  Pred. No. 0.17;
Matches    62;  Conservative    23;  Mismatches    82;  Indels    66;  Gaps    11;

Qy    1  SPHQA--AAPVDQTPRTLATMGQRALPSSLA-----LLSRPLSPPPAACSGDPGCGSGA     52
         |||    |||    ::|||                 |||  ||
Db    8  SPVSAPVAAPV--TPSAVAAPVQVVSPAAVAPAPAAPIAVTPVAPPPTLASVQP------   59

Qy   53  GLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGG  112
         ||       :||               ::
Db   60  ATVTIPAPAPIAAASVTPVASVAPPV---------VAAPTPPAASPVSTPVAVAQIPVA-  109

Qy  113  MASGSSQPWAAASATPMLSSKASLCI-----PTRGPPPQPLMRTPA--------ARS    156

Db  110  -VSAPVAPPVAATPTPVVQIPVAAPVIATPVAASAPTPAAVTPVISPVIASPPVVPANT   168

Qy  157  HWPIPHPCDTACPAPLPVV-----LVAPRSTILSMSRTWTCRRWAVAPCRAE     203
         |||                     ||||
Db  169  TVFVAAPV-AAVPAAVPVVAPVLAPAVAP-----------------AVAPVVAE       204

RESULT 10
S49915
extensin-like protein - maize
C;Species: Zea mays (maize)
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004
C;Accession: S49915
R;Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A;Description: Pex genes: pollen-specific genes with extensin-like domains.
A;Reference number: S49915
```

```
A;Accession: S49915
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1188 <RUB>
A;Cross-references: UNIPROT:Q41805; UNIPARC:UPI0000A1616; EMBL:Z34465; NID:g600117; PID

Query Match    11.4%;  Score    129.5;  DB    2;  Length    1188;
Best Local Similarity    25.8%;  Pred. No. 0.28;
Matches    54;  Conservative    21;  Mismatches    99;  Indels    35;  Gaps    6;

Qy    1  SPHQAAAPVDQTPRTLAT------MGQRALPSSLALLSRP---LSPPPAACSGDPGCGSG   51
         |||                                        ||    ||||||||
Db  507  SPPQ--APVGSPPPVKTTSPPAPIGSPSPSPPPVSVSPPPVKSPPPPAPVGSSPPPPEK    564

Qy   52  AGLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAG  111

Db  565  SPPPPVKS-----SPPPPVPA-----SPPPPVPAPPTLVASPPPVKSPPPAPVASPPPVKSPPPPT  618

Qy  112  GMASGSSQPWAAASATPMLSSKASLCI----PTRGPPPQPLMRTPAARSHWPIPH--    162
         :||                           |   ||||||||||||||||||
Db  619  PVASPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPPPAKSTPPPEEYPTPPTSVK   678

Qy  163  ------PCDTACPAPLPVVLVAPRST    182
               |:|   |||||
Db  679  SSPPPEKSLPPPTLIPSPPPPQEKPTPPST   707

RESULT 11
A40670
nuclear envelope protein POM 121 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C;Accession: A40670
R;Hallberg, E.; Wozniak, R.W.; Blobel, G.
J. Cell Biol. 122, 513-521, 1993
A;Title: An integral membrane protein of the pore membrane domain of the nuclear envelop
A;Reference number: A40670; MUID:93328754; PMID:8335683
A;Accession: A40670
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1199 <HAL>
A;Cross-references: UNIPROT:P52591; UNIPARC:UPI000012FC0E; GB:Z21513; NID:g396746; PIDN:
F:803-807,845-849,956-960,1010-1014,1047-1051,1076-1080/Region: pentapeptide motif (X-F-

Query Match    11.4%;  Score    129.5;  DB    2;  Length    1199;
Best Local Similarity    26.4%;  Pred. No. 0.28;
Matches    55;  Conservative    23;  Mismatches    71;  Indels    59;  Gaps    11;

Qy    1  SPHQAAAPVDQTP--RTLATMGQRALPSSL-ALLSRPL-----SPPPAACSGDPGCGS     50
         |||                    :|                   ||
Db  535  TPATDTSPATSPPPTLTLPTVGPASPASLPAPSSNPLLESLKKMQESPAPSSSEPPEAA   594

Qy   51  GAGLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPA  110

Db  595  TVAAPSPPKTPSLLAPLVSPLTG---------PLAASTSSDSKPTTTFL----       633

Qy  111  GGMASGSSQPWAAASATPMLSSK----ASLCIPT--RGPPPQP----LMRTPAAR     155

Db  634  -GLAS------ASSATPLTDTKAPGVSQAQLCVSTPAATAPSPTPASTLFGMLSPPASS   685

Qy  156  SHWPIPHPCDTACPAPL--PVVLVAPRS     181
         |:|                   ||||
Db  686  SSLATPGP---ACASPMFKPIFPATPKS    710

RESULT 12
QQBE8
BPLF1 protein - human herpesvirus 4 (strain B95-8)
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 09-Jul-2004
C;Accession: G93065; A03747; S32993
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
```

RESULT 13
T43481
probable mucin DKFZp434C196.1 - human (fragment)
N;Alternate names: protein DKFZp434B0635.1
C;Species: Homo sapiens (man)
A;Accession: G86292
A;Molecule type: DNA
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T43481; T34549; T17264
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z22514
A;Accession: T43481
A;Molecule type: mRNA
A;Residues: 1-580 <AAA>
A;Cross-references: UNIPROT:Q9UF83; UNIPARC:UPI000006D400; EMBL:AL133561; NID:g6599133;
R;Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A;Reference number: Z21540
A;Accession: T34549
A;Molecule type: mRNA
A;Residues: 262-580 <POU1>
A;Cross-references: UNIPARC:UPI0000070F36; EMBL:AL220069; NID:g6102864; PIDN:CAB59245.2;
A;Experimental source: adult testis; clone DKFZp434B0635
R;Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A;Reference number: Z18723
A;Accession: T17264
A;Molecule type: mRNA
A;Residues: 262-580 <POU2>
A;Cross-references: UNIPARC:UPI0000070F36; EMBL:AL117481; NID:g5911958; PIDN:CAB55954.1;
A;Experimental source: adult testis; clone DKFZp434B061
C;Genetics:
A;Note: DKFZp434C196.1; DKFZp434B0635.1; DKFZp434B061.1

Query Match   11.4%; Score 129; DB 2; Length 580;
Best Local Similarity 24.6%; Pred. No. 0.16;
Matches 63; Conservative 63; Mismatches 104; Indels 62; Gaps 8;

Qy   1 SPHQAA--APVDQTPRTLATMGQRALPSSLALLSRP----------------------- 34

A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A;Reference number: A93065; MUID:85035713; PMID:6092825
A;Accession: G93065
A;Molecule type: DNA
A;Residues: 1-3149 <BAN>
A;Cross-references: UNIPROT:P03186; UNIPARC:UPI000000CDC7; EMBL:V01555; NID:g59074; PIDN
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A;Reference number: A03794; MUID:84270667; PMID:6087149
A;Contents: annotation; protein coding region
C;Superfamily: human herpesvirus 4 BPLF1 protein

Query Match   11.4%; Score 129.5; DB 1; Length 3149;
Best Local Similarity 27.0%; Pred. No. 0.65;
Matches 53; Conservative 20; Mismatches 80; Indels 43; Gaps 8;

Qy   10 DQTPRTLATMGQRALPSSLALLSRPLSP-----PPAACSGDPGCGSGAGLPSASAAAGIA 64
Db 444 DQLPRTHVPPHR---PPSAARLPPPVIPIPHQSPPASPTHPA-------PVSTIAPSVT 493

Qy   65 SSAVBPV--------CGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASG 116
Db 494 PSPRLPLQIPIPLPQAAPSNPKIPLTTPSPSPTAAAAPTTTLSPPPTQQQPPQSAAPAP 553

Qy 117 S-------SQPWAAASATPMLSSKASLCIPTRGPPQPLMRTPAARSHWPIP---HPCDTAC 168
Db 554 SPLLPQQQPTPSAAPAP-------SPLLPQQQPPPS------AARAPSPLPPQQQQLPSAT 601

Qy 169 PAPLPVVLVAPRSTIL 184
Db 602 PAPPPAAQQLPPSATTL 617

Db 267 SPPRASPMTPPRASPRTPPRASPTTTPSRASLTRTPSWASPTTTPSRASLMKMESTVSIT 326

Qy   35 LSPPPAACSGDPGCGSGAGLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGP 94
Db 327 RTPPRASPTGTPSRASPTGTPSRASLTGSPSRA--SLTGTPSRASLIGTPSRASLIGTPS 384

Qy   95 RSTMECPPALIVHPPAGGMASGSSQPWAAASATP--------------MLSSKASLCIPTRG 142
Db 385 RA------SLTGTPSRASLTGTSST--ASLTRTPSRASLTRTQSSSLTRTPSMASLTRT 436

Qy 143 PPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPRSTI------LSMSRTWTCRRWA 196
Db 437 PPRASLTRTP--------PRASLTRTPPRASLTRTPSRASLTRTPSMVSLKRSPSRASLT 488

Qy 197 VAPCRABKLMCSSSRS 212
Db 489 RTPSRASLTMTPSRAS 504

RESULT 14
G86292
hypothetical protein F7H2.17 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: G86292
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
C;A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1006 <STO>
A;Cross-references: UNIPROT:Q9LMQ1; UNIPARC:UPI00000A2407; GB:AE005172; NID:g8927662; PI
C;Genetics:
A;Map position: 1

Query Match   11.4%; Score 129; DB 2; Length 1006;
Best Local Similarity 28.6%; Pred. No. 0.26;
Matches 63; Conservative 8; Mismatches 93; Indels 56; Gaps 11;

Qy   8 PVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPG-----GAGLPS---- 56
Db 553 PFTMTPPPLLGGGAPGTTDS-------PPPPLLGSSGAPGITGSPPPPLLGGGAPGITGS 604

Qy   57 -----ASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALI----- 105
Db 605 PPPPLLGGGAPGITGSPPPPLLGGGAPG-ITGSPPPPLLGGGAPGITGSPPPPLFGGGAP 663

Qy 106 ----VHPP------AGGMAS---------GSSQPWAAASATPMLSSKASLCIPTRGPPPQPL 148
Db 664 GITGSPPPPLFGGGAPGIAGSPPPPLLGGGAPGITVSPPPLLGGGAP---GITGSPPPPL 720

Qy 149 MR--TPAARSHWPIPHPCDTACPAPLPVVLVAP---RSTI 183
Db 721 VADVPPMPPLAWFSPPDITTGSPPPSPVFLPPPLDRSTL 760

RESULT 15
T44768
antifreeze glycopeptide AFGP polyprotein precursor [imported] - Boreogadus saida
C;Species: Boreogadus saida
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T44768
R;Chen, L.; DeVries, A.L.; Cheng, C.H.C.

Proc. Natl. Acad. Sci. U.S.A. 94, 3817-3822, 1997
A;Title: Convergent evolution of antifreeze glycopeptides in Antarctic nototothenioid fish
A;Reference number: Z22834; MUID:97268653; PMID:9108061
A;Accession: T44768
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-507 <CHE>
A;Cross-references: UNIPROT:O13028; UNIPARC:UPI00000OFD737; EMBL:U43200; NID:g2078482; PI
C;Genetics:
A;Introns: 1/3

```
Query Match        11.3%;   Score 128.5; DB 2; Length 507;
Best Local Similarity  25.6%;   Pred. No. 0.15;
Matches  51; Conservative  23; Mismatches  94; Indels  31; Gaps   5;

Qy      5 AAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPAACSGDPGCGSGAGLPSASAAAGIA 64
          | ||||||:||||||||| |:::||  :|:|| : ||: |:  :|:: : :|:|
Db    282 ATAATDATAATAATPARAATPATPATAATPATPATAATAATAATAATPARAARAAATP 341

Qy     65 SSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAA 124
          ::|:|  :||||  :| | : |:  |:::| |: :| |: :| :::  | ||
Db    342 ATAATP--ATAATAATAATAATAATPARAARAAATPA-----TAATAATAATAAT 394

Qy    125 SATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPA-PLPVVLVAPRSTI 183
          : ||| :::|:   |||    |  |  ||| : |:: ||  | : :| | | :
Db    395 AATPARAARAA---------TPAT------PATPATPATAATAATAATAATA 431

Qy    184 LSMSRTWTCRRWAVAPCRA 202
          : :: ||  ||:| |
Db    432 ATAATAATAATAATAPTPA 450
```

Search completed: March 16, 2006, 15:57:09
Job time : 41 secs

```
          GenCore version 5.1.7
     Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM protein - protein search, using sw model

Run on:   March 16, 2006, 15:49:24 ; Search time 230 Seconds
                                      (without alignments)
                             650.313 Million cell updates/sec

Title:            US-09-989-890-238
Perfect score:    1134
Sequence:   1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:   2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters:        2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  UniProt_05.80:*
            1: uniprot_sprot:*
            2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	775	68.3	207	2	Q68D65_HUMAN	homo sapien
2	680	60.0	216	2	Q92Q00_MOUSE	mus musculu
3	278	24.5	367	2	Q8C1Q5_MOUSE	mus musculu
4	154.5	13.6	1992	2	Q9P6T1_NEUCR	neurospora
5	154	13.6	559	2	Q6ZTM8_HUMAN	homo sapien
6	154	13.6	2427	2	Q6ZRS2_HUMAN	homo sapien
7	154	13.6	3053	2	O15026_HUMAN	homo sapien
8	152	13.4	659	2	Q6C7Q8_YARLI	yarrowia li
9	152	13.4	2971	2	Q9Y5L9_HUMAN	homo sapien
10	150.5	13.3	406	2	Q8VPM8_9MICC	micrococcus
11	143	12.6	780	2	Q6CB04_YARLI	yarrowia li
12	139	12.3	609	2	Q8FM53_COREF	corynebacte
13	138.5	12.2	829	2	Q6N5Q1_RHOPA	rhodopseudo
14	138.5	12.2	1514	2	Q4FX85_LEIMA	leishmania
15	138	12.2	546	2	Q4PF94_USTMA	ustilago ma
16	137	12.1	269	2	Q6I5V3_ORYSA	oryza sativ
17	137	12.1	616	2	Q9H6K5_HUMAN	homo sapien
18	136.5	12.0	635	2	Q9RW96_DEIRA	deinococcus
19	135.5	11.9	463	2	Q63VC5_BURPS	burkholderi
20	135.5	11.9	803	2	Q9P682_NEUCR	neurospora
21	135.5	11.9	817	2	Q7RY24_NEUCR	neurospora
22	135.5	11.9	864	2	CPN_DROME	drosophila
23	135	11.9	3164	1	TEGU_HHV11	human herpe
24	134.5	11.9	328	2	Q6TVP0_9POXV	orf virus.
25	134.5	11.9	1481	2	Q5THJ1_HUMAN	homo sapien
26	134.5	11.9	1540	2	Q5THJ0_HUMAN	homo sapien
27	134.5	11.9	1576	2	Q59H82_HUMAN	homo sapien
28	134.5	11.9	1718	1	PRDM2_HUMAN	homo sapien
29	134	11.8	475	2	Q5Z2I3_NOCFA	nocardia fa
30	133	11.7	1325	2	Q9BKV7_LEIMA	leishmania
31	133	11.7	1435	2	Q4FX64_LEIMA	leishmania
32	133	11.7	2408	2	Q60DH6_ORYSA	oryza sativ
33	133	11.7	7194	2	Q4FX63_LEIMA	leishmania
34	132.5	11.7	383	2	Q4LN41_9BURK	burkholderi
35	132.5	11.7	850	2	Q6K322_ORYSA	oryza sativ
36	132.5	11.7	969	2	Q4VBD6_MOUSE	mus musculu
37	132.5	11.7	1165	2	Q42TJ8_TETNG	tetraodon n
38	132.5	11.7	3084	2	Q5PPB8_9ALPH	suid herpes
39	132.5	11.6	3084	2	Q8UZ11_PRVKA	pseudorabie
40	132	11.6	243	2	Q9XIV1_CUCSA	cucumis sat
41	132	11.6	574	1	WSP1_SCHPO	schizosacch
42	132	11.6	3326	1	Q7T591_CHV1	cercopithec
43	131.5	11.6	190	2	Q6KIT4_ORYSA	oryza sativ
44	131.5	11.6	17392	2	Q4FX62_LEIMA	leishmania
45	130.5	11.5	540	2	Q65OZ7_ORYSA	oryza sativ

ALIGNMENTS

```
RESULT 1
Q68D65_HUMAN
ID   Q68D65_HUMAN PRELIMINARY;            PRT;    207 AA.
AC   Q68D65;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein DKFZp686B17277.
GN   Name=DKFZp686B17277;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Prostate;
RG   The German cDNA Consortium;
RA   Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (AUG-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR749558; CAH18355.1; -; mRNA.
KW   Hypothetical protein.
SQ   SEQUENCE   207 AA;  22216 MW;  FC6B65C436F28BD7 CRC64;

    Query Match             68.3%;
    Best Local Similarity   72.6%;  Score 775;  Pred. No. 9.3e-41;
    Matches  156;  Conservative   10;  Mismatches   31;  Indels   18;  Gaps    5;

Qy    2 PHQAAAPVDQTPRTLATMGQRALPSSLAL--LSRPLSPPPAACS--GDPGCGSGAGLPSA 57
          | ||  :  ||    ::  |  :: ||  :  |||:  |||
Db    7 PRPASGP----PR----LGPWERPTELCLETYDKPQPPSRRTRRPDPKDPGHHGPESI 58

Qy   58 SAAAGIASSAAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGS 117
                             :|  ||||||||||||||||||||||||||||||||||
Db   59 TFISGSAEPALE------SPTCLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGS 112

Qy  118 SQPWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLV 177
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  113 SQPWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLV 172

Qy  178 APRSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 212
         |||||||||||||||||||||||||||||||||||
Db  173 APRSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 207


RESULT 2
Q92Q00_MOUSE
ID   Q92Q00_MOUSE PRELIMINARY;            PRT;    216 AA.
AC   Q92Q00;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   RIKEN cDNA 1810019J16.
```

GN   Name=1810019J16Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.F.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC006890; AAH06890.1; -; mRNA.
DR   MGI; MGI:1916323; 1810019J16Rik.
SQ   SEQUENCE 216 AA; 21683 MW; F7AD1985BC0976D8 CRC64;

Query Match   60.0%; Score 680; DB 2; Length 216;
Best Local Similarity   67.5%; Pred. No. 7.5e-35;
Matches 143; Conservative   8; Mismatches   61; Indels   0; Gaps   0;

Qy   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCCGSGAGLPSASAA   60
          ||| |||||| || | |||||:||||| || |||: | || ||||| ||||||||
Db   5 SPHRAVAPGGQTLRTLATMGQRVSPSFQALQNQPTSPQPAAASSGAPGVGTGVGLPSASAA   64

Qy   61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGWASGSSQP 120
          ||||||||||:||| |||| ||||| ||||| ||||| |||| ||||| || |||| |
Db   65 AGIACSAVGLVCGAAAPAYLPETPLKGLPKPPGPRNTMVCPPARTVHPPAAGMARGSRPA 124

Qy   121 WAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR 180
          ||||||||||||||| ||| ||||||||||||| |||||| | | |||| |||||
Db   125 WAAASATLMLSSKASLSTPTAMPPPQSLTWTPAARSPWPSLLPHGTACLAPSPTAPAALR 184

Qy   181 STILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 212
          ||:|:||| |||||| ||||| |||| |||
Db   185 STTPSMNRTWTCLRWAVAPCRAGRSTCLFSRS 216

RESULT 3
Q8C1Q5_MOUSE
ID   Q8C1Q5 MOUSE PRELIMINARY;   PRT;   367 AA.
AC   Q8C1Q5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched
DE   library, clone:1810019J16 product:SIMILAR TO TATA BOX BINDING PROTEIN
DE   (TBP)-ASSOCIATED FACTOR, RNA POLYMERASE II, C1, 130KD homolog.
GN   Name=1810019J16Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Pancreas;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Pancreas;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Pancreas;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Pancreas;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Pancreas;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Pancreas;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato M.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,

RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata Y., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK007551; BAC25177.1; -; mRNA.
DR   Ensembl; ENSMUSG00000037600; Mus musculus.
DR   MGI; MGI:1916323; 1810019J16Rik.
SQ   SEQUENCE   367 AA;   38538 MW; D558AD91D04FF473 CRC64;

Query Match   24.5%; Score 278; DB 2; Length 367;
Best Local Similarity   40.8%; Pred. No. 9.9e-10;
Matches   78; Conservative   17; Mismatches   78; Indels   18; Gaps   4;

Qy   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60
                |||||:|| ||| ||||||| :|:|||:|:|||||| ||| || |||||||||||||| :
Db   5 SPHRAVAPGGQTLRTLATTQRVSPSFQALQNQPTSPQPAASSGAPGVGTGVGLPSASDA 64

Qy   61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP 120
                |||:||| |||| ||||| :||| :| ||| | |:| |:
Db   65 AGIACSAVGLVCGAAAPALSAGDPIEGSARAAWAKEHNGVPPSPDRAPP------SRR 116

Qy   121 WAAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPITPHPCDTACPAPLPVVLVAPR 180
                       |: :| ||| | :|| :|| ||| ||||| |
Db   117 DGQRLKTSMGSSFSYPDVKLKGIPVVPYRHATS-----PVP-DVDSCCKEP----LAEPP 166

Qy   181 STILSMSRTWT 191
                 | : |: |
Db   167 PTRHSLPSTFT 177

RESULT 4
Q9P6T1_NEUCR
ID   Q9P6T1_NEUCR   PRELIMINARY;   PRT;   1992 AA.
AC   Q9P6T1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein 15B6.220.
GN   Name=15B6.220;
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RA   German Neurospora genome project;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353822; CAB88653.2; -; Genomic_DNA.
DR   PIR; T48814; T48814.
     HSSP; Q63244; 1KQ8.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000637; A+T_hook.
DR   InterPro; IPR001766; TF Fork_head.
DR   InterPro; IPR011991; Wing_hlX_DNA_bd.
DR   Pfam; PF02178; AT_hook; 2.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00929; ATHOOK.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM00384; AT_hook; 2.
DR   SMART; SM00339; FH; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
KW   Hypothetical protein.

SQ   SEQUENCE   1992 AA;   207430 MW; CA84145FD154281B CRC64;

Query Match   13.6%; Score 154.5; DB 2; Length 1992;
Best Local Similarity   30.7%; Pred. No. 0.2;
Matches   62; Conservative   20; Mismatches   77; Indels   43; Gaps   9;

Qy   1 SPHQAAAPVDQT-PRTLATMGQRALP---SSLALLSRPLSPPPAACSGDPGCGSGAGLP 55
                 |::| |::| |::| | |:| | :|:|||| ||| ||
Db   1801 TPHSASRPVPQSVPQSVPQATPQAVPRPSTSALTPTAQPGPVSPAV-------SGSGVP 1852

Qy   56 SASAAAGIASSAVE--PV-CGDAAPACLLR---TPLRGLLKP------TGPRSTMECP- 101
                 |:|| :|:|: | || | ||||| ||:   :| :|| TGPRSTMECP
Db   1853 APSAAQSVAPAPVSSTPVPAATVAAPTPRTVTAAPALSAATNPAPVPSQPQH 1912

Qy   102 ------PALIVHPPAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGPPQPLMRTPAAR 155
                      PALIVHP   QAQAPATPTITSAAPPRPPTLAPPPPPPPTPED
Db   1913 QITGQAPAQQQRPPA---------------QAQAPATPTITSAAPPRPPTLAPPPPPPPTPED 1962

Qy   156 SHWPIPHPCDTACPAPLPVVLV 177
                                :
Db   1963 PPPPPPPPPAEAPPPPPPTPLM 1984

RESULT 5
Q6ZTM8_HUMAN
ID   Q6ZTM8_HUMAN   PRELIMINARY;   PRT;   559 AA.
AC   Q6ZTM8;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ44499.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Uterus;
RA   Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA   Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK126463; BAC86558.1; -; mRNA.
SQ   SEQUENCE   559 AA;   54028 MW; B5E50A2A892BA942 CRC64;

Query Match   13.6%; Score 154; DB 2; Length 559;
Best Local Similarity   25.7%; Pred. No. 0.07;
Matches   69; Conservative   27; Mismatches   92; Indels   80; Gaps   12;

Qy   2 PHQAAAPVD---QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS 58
                 |:|:| :|::|| : | |: |: | ||:   :| || |||||||:|
Db   97 PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVVPAAPGPPSLAPSGASPSAS 152

Qy   59 A-AAGIA----------------------SSAVEPVCGDA-APACLLRTPLRGLL 89
                 | |:| SSAVEPVCGDA APACLLRTPLRGLL
Db   153 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 212

Qy   90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGPP----- 144
                                       ||:| | ::|:|| || :
Db   213 NPAPAQASLLAP------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 259

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT--------ACPAPLPVVLVAPRST------ 182
                       :| | :| | : : :| ACPAPLPVVLVAPRST
Db   260 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 319

Qy   183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
                 :| :|:| : | | :| ::|
Db   320 FLPSPASTQTL---ALAPALAPTLGSS 344

RESULT 6
Q6ZRS2_HUMAN
ID   Q6ZRS2_HUMAN PRELIMINARY;   PRT; 2427 AA.
AC   Q6ZRS2;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ46149.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Testis;
RA   Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda A., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR   EMBL; AK128030; BAC87237.1; -; mRNA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0004386; F:helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR006562; HSA.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; Helicase_C; 1.
DR   Pfam; PF07529; HSA; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00573; HSA; 1.
KW   DNA damage; DNA repair; DNA-binding; Helicase; Hydrolase;
KW   Nuclear protein.
FT   NON_TER   2427   2427
SQ   SEQUENCE   2427 AA; 260573 MW; 1BD1EDE747FF5927 CRC64;

Query Match   13.6%; Score 154; DB 2; Pred. No. 0.26;
Best Local Similarity   25.7%;
Matches   69; Conservative   27; Mismatches   92; Indels   80; Gaps   12;

Qy   2 PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS 58
            |:||:|| |:|:| :: :: :||:::|||:| ||||
Db   1448 PAPASAPLTIPISAPLTVSASG---PALLTSVTPPLAPVVPAAPGPPSLAPSGASPSAS 1503

Qy   59 A-AAGIA-----------------------SSAVEPVCGDA-APACLLRTPLRGLL 89
            | :|| ||| |||:| :| ::|| || |:
Db   1504 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1563

Qy   90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP----- 144
            ||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db   1564 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP----- 1610

Qy   145 NPAPAQASLLAP-----------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 182
                 : :|| |::||:|| || | || || || | ||| || |
Db   1611 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1670

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT-------ACPAPLPVVLVAPRST-------- 182
               : |:: : || || | :|| | :| || | |: |||
Db   1611 ILSMSRTWTCRRWAVAPCRAEKLMCSS 209

Qy   183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
             ||||||||||||||||||||||||||||
Db   1671 PLPSPASTQTL---ALAPALAPTLGGSS 1695

RESULT 7
O15026_HUMAN
ID   O15026_HUMAN PRELIMINARY;   PRT; 3053 AA.
AC   O15026;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   KIAA0309 protein (Fragment).
GN   Name=KIAA0309;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
DR   EMBL; AB002307; BAA20768.2; -; mRNA.
DR   Ensembl; ENSG00000080603; Homo sapiens.
DR   GO; GO:0004402; F:histone acetyltransferase activity; TAS.
DR   GO; GO:0003713; F:transcription coactivator activity; TAS.
DR   GO; GO:0006357; P:regulation of transcription from RNA polyme. . .; TAS.
DR   InterPro; IPR000637; A+T hook.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR006562; HSA.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF02178; AT_hook; 3.
DR   Pfam; PF00271; Helicase_C; 1.
DR   Pfam; PF07529; HSA; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   PRINTS; PR00929; ATHOOK.
DR   SMART; SM00384; AT_hook; 3.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00573; HSA; 1.
FT   NON_TER   1   1
SQ   SEQUENCE   3053 AA; 325778 MW; 721B4861D37E19C2 CRC64;

Query Match   13.6%; Score 154; DB 2; Length 3053;
Best Local Similarity   25.7%; Pred. No. 0.32;
Matches   69; Conservative   27; Mismatches   92; Indels   80; Gaps   12;

Qy   2 PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS 58
            |:||:|| |:|:| :: :: :||:::|||:| ||||
Db   1271 PAPASAPLTIPISAPLTVSASG---PALLTSVTPPLAPVVPAAPGPPSLAPSGASPSAS 1326

Qy   59 A-AAGIA-----------------------SSAVEPVCGDA-APACLLRTPLRGLL 89
            | :|| ||| |||:| :| ::|| || |:
Db   1327 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1386

Qy   90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP----- 144
            ||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db   1387 NPAPAQASLLAP-----------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1433

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT-------ACPAPLPVVLVAPRST-------- 182
                 : |:: : || || | :|| | :| || | |: |||
Db   1434 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1493

Qy   183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
             ||||||||||||||||||||||||||||
Db   1494 PLPSPASTQTL---ALAPALAPTLGGSS 1518

RESULT 8
Q6C7Q8_YARLI

```
ID   Q6C7Q8_YARLI PRELIMINARY;      PRT;   659 AA.
AC   Q6C7Q8;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to tr|Q95JC9 Sus scrofa Basic proline-rich protein.
GN   OrderedLocusNames=YALI0D26191g;
OS   Yarrowia lipolytica (Candida lipolytica).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Yarrowia.
OX   NCBI_TaxID=4952;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX   PubMed=15229592; DOI=10.1038/nature02579;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA   Boisramé A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA   Swennen D., Tekaia F., Wésolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.-L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
DR   EMBL; CR382130; CAG81510.1; -; Genomic_DNA.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 1.
KW   Complete proteome.
SQ   SEQUENCE   659 AA;   62466 MW;   2CE5088C9EF0908 CRC64;

Query Match             13.4%;  Score 152;  DB 2;  Length 659;
Best Local Similarity   27.4%;  Pred. No. 0.11;
Matches   57;  Conservative  21;  Mismatches  88;  Indels  42;  Gaps   9;

Qy      1  SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGGAGLPSASAA   60
Db    181  SPPQQGAPA--IPSLRKTSGAPSAPGGFA-----PPPPAPPGAPAIPGAPSVASSYRS   233

Qy     61  -VBPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGM   113
Db    234  ASASSGAPPPPGGAPIPGAPP-----PLPGKVSTSGGAPTFGAPP---PPPGGAP   284

Qy    114  ASGSSQPWAAASATPMLSSKASLCIPTRGPP------HPC   164
Db    285  AYGAPAPPTPGTSSPKPPPK-----PAKRPPALKPKPIPTPGLKPAVPTPGQRRSVSPS   339

Qy    165  DTACPAPL----PVVLVAPRSTILSMS   187
Db    340  PGAPPPPIPGSVAPSVRHAPSQSVSSIA   367
```

```
RESULT 9
Q9YSL9_HUMAN
ID   Q9YSL9_HUMAN PRELIMINARY;      PRT;   2971 AA.
AC   Q9YSL9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Transcriptional activator SRCAP.
GN   Name=SRCAP;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=99278407; PubMed=10347196; DOI=10.1074/jbc.274.23.16370;
RA   Johnston H., Kneer J., Chackalaparampil I., Yaciuk P., Chrivia J.;
RT   "Identification of a novel SNF2/SWI2 protein family member, SRCAP,
RT   which interacts with CREB-binding protein.";
RL   J. Biol. Chem. 274:16370-16376(1999).
DR   EMBL; AF143946; AAD39760.1; -; mRNA.
DR   TRANSFAC; T04151; -.
DR   Ensembl; ENSG00000080603; Homo sapiens.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0004386; F:helicase activity; IEA.
DR   InterPro; IPR000637; A+T_hook.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF02178; AT_hook; 3.
DR   Pfam; PF00271; Helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   PRINTS; PR00929; ATHOOK.
DR   SMART; SM00384; AT_hook; 3.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
SQ   SEQUENCE   2971 AA;  315641 MW;  F9F7EB70304B78A3 CRC64;

Query Match             13.4%;  Score 152;  DB 2;  Length 2971;
Best Local Similarity   25.7%;  Pred. No. 0.41;
Matches   69;  Conservative  27;  Mismatches  92;  Indels  80;  Gaps  12;

Qy      2  PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGGAGLPSAS   58
Db   1189  PAPASAPLTIPISAPLTVSASG----PALLTSVTPLAPVVPAAPGPPSLQPSGASPSAS  1244

Qy     59  A-AAGIA------------------SSAVBPVCGDA-APACLLRTPLRGLL   89
Db   1245  ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP  1304

Qy     90  KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP-----  144
Db   1305  NPAPAQASLLAP--------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP  1351

Qy    145  ---PQPLMRTPAARSHWPIP--HPCDT-------ACPAPLPVVLVAPRST---------  182
Db   1352  VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQMLPAPVPS  1411

Qy    183  -ILSMSRTWTCRRWAVAPCRAEKLMCSS  209
Db   1412  PLPSPASTQTL--ALAPALAPTLGGSS  1436
```

```
RESULT 10
Q8VPM8_9MICC
ID   Q8VPM8_9MICC PRELIMINARY;      PRT;   406 AA.
AC   Q8VPM8;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Proline-rich extensin-like protein.
OS   Micrococcus sp. 28.
OG   Plasmid pSD10.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Micrococcineae; Micrococcaceae; Micrococcus.
OX   NCBI_TaxID=161213;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=28;
RA   Zhong Z., Caspi R., Mincer T., Helinski D., Knauf V., Boardman K.,
RA   Wilkinson J.E., Shea T., DeLoughery C., Toukdarian A.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY034092; AAK62519.1; -; Genomic_DNA.
KW   Plasmid.
SQ   SEQUENCE   406 AA;   38781 MW;   2E23C7D5637398B CRC64;

Query Match             13.3%;  Score 150.5;  DB 2;  Length 406;
```

```
Best Local Similarity  27.7%;  Pred. No. 0.087;
Matches  62;  Conservative  18;  Mismatches  101;  Indels  43;  Gaps  7;

Qy    1 SPHQAAAPVD----QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPS 56
          |||    :||||
Db   18 SPAAGAAPPDSSGVSTPPDASPAGASSLPEVPAAPPAEDAPAPSPEPDSPPRPETAPAPE 77

Qy   57 ASAAAGIASSAVEPVCGDAA--PACLLRTPLR-GLLKPTGPRSTMECP----------- 101
          :|:|.||:          |:.    |:
Db   78 ATALLGDRAAIAPAPAGRSAAFPSLAAAEPLAPPATAPPWPTPVWEAPSPDLSPEFRGRT 137

Qy  102 ---PALIVHPPAGGMASGSSQPWA------------AASATPMLSSKASLC-----IPT 140
             |||||||||||||||||::|
Db  138 PGAAALADWPPAAGATPAAARPAAVPARAPLIGDAPPAAGLTPAPESSPSLASTRSTVPS 197

Qy  141 RGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPRSTIL 184
          |::|||| |||| ||
Db  198 TPPAPVPRAAEVPPAA---PVPWSENSAEDAPLPAAPAAPLSALL 238
```

RESULT 11
Q6CB04_YARLI
```
ID   Q6CB04_YARLI  PRELIMINARY;    PRT;    780 AA.
AC   Q6CB04;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to sp|P08640 Saccharomyces cerevisiae YIR019c STA1
DE   extracellular alpha-1.
GN   OrderedLocusNames=YALI0C22924g;
OS   Yarrowia lipolytica (Candida lipolytica).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Yarrowia.
OX   NCBI_TaxID=4952;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX   PubMed=15229592; DOI=10.1038/nature02579;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA   Swennen D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.-L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
DR   EMBL; CR382129; CAG82478.1; -; Genomic_DNA.
KW   Complete proteome.
SQ   SEQUENCE  780 AA;  73395 MW;  F6AABB2C4043A7E9 CRC64;
```

```
Qy  180 RST 182
          |:|
Db  254 ASS 256
```

RESULT 12
Q8FM53_COREF
```
ID   Q8FM53_COREF  PRELIMINARY;    PRT;    609 AA.
AC   Q8FM53;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=CE2654;
OS   Corynebacterium efficiens.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=152794;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RX   MEDLINE=22723752; PubMed=12840036; DOI=10.1101/gr.1285603;
RA   Nishio Y., Nakamura Y., Kawarabayasi Y., Usuda Y., Kimura E.,
RA   Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,
RA   Gojobori T.;
RT   "Comparative complete genome sequence analysis of the amino acid
RT   replacements responsible for the thermostability of Corynebacterium
RT   efficiens.";
RL   Genome Res. 13:1572-1579(2003).
DR   EMBL; BA000035; BAC19464.1; -; Genomic_DNA.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE  609 AA;  61385 MW;  D127080D3874A578 CRC64;
```

```
Query Match  12.3%;  Score 139;  DB 2;  Length 609;
Best Local Similarity  29.1%;  Pred. No. 0.64;
Matches  72;  Conservative  18;  Mismatches  103;  Indels  54;  Gaps  12;

Qy    5 AAAPVDQTPRTLATMGQ--------------RALPSSLALLSRPLSPPPAACSGD----PGCGS 50
          |||                           ||||...||||.:|||:||
Db   63 AAVPAPATPIPPAAPGSAIPAPGAVTPTAVPTPGAAVPATPTPAAPGSAIPAPGAAT 122

Qy   51 ------------GAGLPGSAAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKP--TGP 94
                      |.:|          ||.|
Db  123 PTAVPTPGSAIPTGAAVPAPGVATPSAPGASIPTPGAAMPTPGTATPAPGAAAPGATIP 182

Qy   95 RSTMECP--PALIVHPPAGGMAS------GSSQPWAAA--SATPMLSSKASLCIPTRGP 143
          |:.||     |||:||||              ||:        ||.:::
Db  183 GSAVPAPGGAPAAPGAPAAPRTPAPGAAPRTPAPGAVPGAVPGSAVP---TPGAISAPGAAP 239

Qy  144 PP-QPLMRTPAARSHWPIP---HPCDTACPAPLPVVLVAPRSTILSMSRTWTCRRWAVA 198
          ||                    |||||:||||||||.|       .:  ::
Db  240 PPGLPAPGPGAPGAPGIPAAPGAPGSAAAPAPTSV----PRSAAPVAADTDTRPKGSTA 295

Qy  199 PCRAEKL 205
          |
Db  296 -LRVEEL 301
```

RESULT 13
Q6NSQ1_RHOPA
```
ID   Q6NSQ1_RHOPA  PRELIMINARY;    PRT;    829 AA.
AC   Q6NSQ1;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein precursor.
GN   OrderedLocusNames=RPA2923;
OS   Rhodopseudomonas palustris.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Rhodopseudomonas.
OX   NCBI_TaxID=1076;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
```

```
Query Match  12.6%;  Score 143;  DB 2;  Length 780;
Best Local Similarity  30.1%;  Pred. No. 0.45;
Matches  55;  Conservative  23;  Mismatches  95;  Indels  10;  Gaps  5;

Qy    2 PHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSG--AGLPSASA 59
          . :|||     :||||||||||||||||||||||||||||||:|||.
Db   82 PSSSAAPVSSVPSSSAAPVSSSVPSSSAAPVSSSVPSSSAAPVSSVPSSSA 141

Qy   60 AAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQ 119
          ||||||||||    || |
Db  142 AAGSASEA--PVAANSTSPVASSAP----VSSTTPVSS-TTPVSSVAPSSAVPVASNSTS 194

Qy  120 PWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAP 179
          || :|: ||| | :
Db  195 P-ASSAAPVSSTTPVSSAAPSSEAPVAANSTTPAASSAAPVSSCATTPLPSSAPANSTAP 253
```

RC   STRAIN=CGA009 / ATCC BAA-98;
RX   PubMed=14704707; DOI=10.1038/nbt923;
RA   Larimer F.W., Chain P., Lamerdin J.E., Malfatti S., Do L.,
RA   Land M.L., Pelletier D.A., Beatty J.T., Lang A.S., Tabita F.R.,
RA   Gibson J.L., Hanson T.E., Bobst C., Torres y Torres J.L., Peres C.,
RA   Harrison F.H., Gibson J., Harwood C.S.;
RT   "Complete genome sequence of the metabolically versatile
RT   photosynthetic bacterium Rhodopseudomonas palustris.";
RL   Nat. Biotechnol. 22:55-61(2004).
DR   EMBL; BX572602; CAE28364.1; -; Genomic_DNA.
DR   GO; GO:0030693; F:caspase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR011600; Pept_C14_p20.
DR   Pfam; PF00656; Peptidase_C14; 1.
KW   Complete proteome; Hypothetical protein; Signal.
FT   SIGNAL   1   28   Potential.
SQ   SEQUENCE   829 AA; 85716 MW; B700FD96D06B4EA8 CRC64;

Query Match   12.2%; Score 138.5; DB 2; Length 829;
Best Local Similarity   29.7%; Pred. No. 0.91;
Matches   62; Conservative 14; Mismatches 86; Indels 47; Gaps 13;

Qy   7 APVDQTP--RTLATMGQRALPSSLALLSRP----------LSP--PPAACSGD-- 45
                | ||| || |||||||:||::||| | | |||:|||
Db   618 APAGAAPAVQALPVPGTKGLPPAPGVAARPGIPSVAQPQPPGRPALGPGPGAAARNGTVA 677

Qy   46 PGCGSG----AGLPSASAAAGIASSAVEPVCGDAAPACLLR--TPLRGLLKPTGPRSTME 99
                     ||| | |||| |||:| || |:|| |:| |||||||||||||||
Db   678 PSAGSAPKPLAGTPPAGGPAVRPEAVRP--QQQAPAARLRPTPPVTAPARPAGP----- 730

Qy   100 CPPALIVH---PPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPPPQ---PLMRTPA 153
                |||||: || | ::| ||||| | | ::|| |||||||| |:| |
Db   731 -PPAAAVDRRPPRAAPRIQRPAPPTVSRPVPPPMHVAPRVAPP---PPPDQHAAPRMAPPP 786

Qy   154 A--RSHWPIPHPCDTACPAPLPVVLVAPR 180
                  |:| ||| | | |||||||||||||
Db   787 APVRAAPPPPHVAPPRPAP-PRAAPPPR 814

RESULT 14
Q4FX85_LEIMA
ID   Q4FX85_LEIMA PRELIMINARY;   PRT;   1514 AA.
AC   Q4FX85;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=LMJ_0963;
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Friedlin;
RX   PubMed=16020728; DOI=10.1126/science.1112680;
RA   Ivens A.C., Peacock C.S., Worthey E.A., Murphy L., Aggarwal G.,
RA   Berriman M., Sisk E., Rajandream M.A., Adlem E., Aert R., Anupama A.,
RA   Apostolou Z., Attipoe P., Bason N., Bauser C., Beck A., Beverley S.M.,
RA   Bianchettin G., Borzym K., Bothe G., Bruschi C.V., Collins M.,
RA   Cadag E., Ciarloni L., Clayton C., Coulson R.M., Cronin A., Cruz A.K.,
RA   Davies R.M., De Gaudenzi J., Dobson D.E., Duesterhoeft A.,
RA   Fazelina G., Fosker N., Frasch A.C., Fraser A., Fuchs M., Gabel C.,
RA   Goble A., Goffeau A., Harris D., Hertz-Fowler C., Hilbert H., Horn D.,
RA   Huang Y., Klages S., Knights A., Kube M., Larke N., Litvin L.,
RA   Lord A., Louie T., Marra M., Masuy D., Matthews K., Michaeli S.,
RA   Mottram J.C., Muller-Auer S., Munden H., Nelson S., Norbertczak H.,
RA   Oliver K., O'Neil S., Pentony M., Pohl T.M., Price C., Purnelle B.,
RA   Quail M.A., Rabbinowitsch E., Reinhardt R., Rieger M., Rinta J.,
RA   Robben J., Robertson L., Ruiz J.C., Rutter S., Saunders D.,
RA   Schafer M., Schein J., Schwartz D.C., Seeger K., Seyler A., Sharp S.,
RA   Shin H., Sivam D., Squares R., Squares S., Tosato V., Vogt C.,
RA   Volckaert G., Wambutt R., Warren T., Wedler H., Woodward J., Zhou S.,
RA   Zimmermann W., Smith D.F., Blackwell J.M., Stuart K.D., Barrell B.,

RA   Myler P.J.;
RT   "The Genome of the Kinetoplastid Parasite, Leishmania major.";
RL   Science 309:436-442(2005).
DR   EMBL; CP000081; AAZ14258.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   1514 AA; 150839 MW; 88663AD628129720 CRC64;

Query Match   12.2%; Score 138.5; DB 2; Length 1514;
Best Local Similarity   27.7%; Pred. No. 1.6;
Matches   64; Conservative 23; Mismatches 111; Indels 33; Gaps 9;

Qy   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSP----PPAACSGDPGCGSGAGLP 55
                 :||| |:| || |||||||||||||||||:| ||| | ||| ||
Db   293 APAPAPAAPKAAPAAPAAPAAPKAAPAAPAAPAAPKAAPAAPAAPAAPKAAPAAPAAP 352

Qy   56 SASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECP--PALIVHPPAGGM 113
                ||:|| | ||| ::| | ||||| |:| |:||| || :||| ||
Db   353 KAAPAAPAAPAA--PAAPKAAPAA------------------PAAPKAAPAAPAAPKAAPAAPAPA 399

Qy   114 ASGSSQPWAAASATPM---LSSKASLCIPTRGPPPQPLMRTPAARSHWPIP-----HPCD 165
                | | | |||| | ||||||||||||| ||||||| | | | | ||
Db   400 APKAAPAAPAAPKAAPAAPKAAPAAPAAPAAPKAAPAAPAAPKAAPVPPAAPAAPAA 459

Qy   166 TACPAPLPVVLVAPRSTILSMSRTW--TCRRWAVA--PCRAEKLMCSSSRS 212
                | | ||||||||:|: |: | | :||| || | |:| ||:
Db   460 PAAPKAAPVPPAAP-SVLSAPAVFWRVQSRKGSVAGGSYAARQRMSSSGSS 509

RESULT 15
Q4PF94_USTMA
ID   Q4PF94_USTMA PRELIMINARY;   PRT;   546 AA.
AC   Q4PF94;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=UM01219.1;
OS   Ustilago maydis 521.
OC   Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC   Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX   NCBI_TaxID=237631;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=521;
RA   Birren B., Nusbaum C., Abebe A., Abouelleil A., Adekoya B.,
RA   Ait-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA   Arachchi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA   Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavsky L.,
RA   Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA   Calvo S., Camarata J., Campo A., Chang J., Cheshatsang Y., Citroen M.,
RA   Collymore A., Considine T., Cook P., Cook P., Corum B., Cuomo C.,
RA   David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA   Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA   Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA   Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
RA   Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA   Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA   Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby B., Iliev I.,
RA   Jaffe D., Jones C., Kamal M., Kamat A., Kamvysselis M., Karlsson E.,
RA   Kells C., Kieu A., Kisner P., Kodira C., Kulbokas E., Labutti K.,
RA   Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA   Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA   Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean J., Major J.,
RA   Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA   Mccarthy M., Mcdonough S., Mcghee T., Meldrim J., Meneus E.,
RA   Mesirov J., Mihalev A., Mihova T., Mikkelsen T., Mlenga V., Moru K.,
RA   Mozes J., Mulrain L., Munson G., Naylor J., Newes C., Nguyen C.,
RA   Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA   Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,
RA   O'neill K., Osman S., Parker S., Perrin D., Phunkhang P., Piqani B.,
RA   Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA   Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA   Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,

RA   Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA   Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA   Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA   Tenzing P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA   Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,
RA   Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA   Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA   Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA   Zimmer A., Zody M., Lander E.;
RT   "The genome sequence of Ustilago maydis.";
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC   EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC   preliminary data.
DR   EMBL; AACP01000043; EAK81826.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   546 AA; 55478 MW; 57695SABFAC759AA CRC64;

Query Match   12.2%; Score 138; DB 2; Length 546;
Best Local Similarity   30.4%; Pred. No. 0.67;
Matches   59; Conservative   19; Mismatches   70; Indels   46; Gaps   9;

Qy   5 AAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAAAGIA 64
            ||| | | |: || |: || | || | | : | :: |||
Db   112 AAAEAAQQQATVA-----AIPS------PASQPPAKKRGRP---SKASLEAAAAAA-A 155

Qy   65 SSAVEP----VCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP 120
            :: : | ::| |: : | : || || :   : | | |:: :
Db   156 AAPVAPNGVGISAAAAPPSSIQKKPGR------PPKNTAAQPPAPVVVAPAALPAAAPSAP 209

Qy   121 WAAASATPMLSSKASLC--------IPTRGPPPQ-------PLMRTPAARSHWPIPHPCD 165
              : | |: | | | ||: | | | : |
Db   210 ATKKRGRPPKNVEASFAAAAIPAAPAPTAAEAPQLKKRGRPPKAQTATAQPLVPVQQPF- 268

Qy   166 TACPAPLPVVLVAP 179
            || ||
Db   269 ----APAPVAPTAP 278

Search completed: March 16, 2006, 15:56:24
Job time : 233 secs

```
                    GenCore version 5.1.7
             Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on:    March 16, 2006, 15:56:40 ; Search time 48 Seconds
                                       (without alignments)
                              365.151 Million cell updates/sec

Title:         US-09-989-890-238
Perfect score: 1134
Sequence:      1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters:     572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued Patents AA:*
            1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
            2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
            3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
            4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
            5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
            6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	152	13.4	2972	2	US-09-579-181-2	Sequence 2, Appli
2	152	13.4	3118	2	US-09-579-181-1	Sequence 1, Appli
3	134.5	11.9	308	2	US-09-252-991A-18097	Sequence 18097, A
4	134.5	11.9	1540	2	US-09-949-016-11382	Sequence 11382, A
5	134.5	11.9	1540	2	US-09-949-016-11383	Sequence 11383, A
6	134.5	11.9	1719	1	US-08-455-568-4	Sequence 4, Appli
7	134.5	11.9	1719	1	US-08-399-411-4	Sequence 4, Appli
8	134.5	11.9	1719	2	US-08-516-859A-4	Sequence 4, Appli
9	134.5	11.9	1719	2	US-09-586-472-4	Sequence 4, Appli
10	134.5	11.9	1719	2	US-09-528-706-4	Sequence 4, Appli
11	134.5	11.9	1719	2	US-10-024-450-4	Sequence 4, Appli
12	133.5	11.9	339	2	US-09-252-991A-28087	Sequence 28087, A
13	128	11.3	710	2	US-10-104-047-3402	Sequence 3402, Ap
14	127	11.2	246	2	US-10-104-047-2232	Sequence 2232, Ap
15	127	11.2	246	2	US-10-104-047-2243	Sequence 2243, Ap
16	127	11.0	369	2	US-09-252-991A-25394	Sequence 25394, A
17	124.5	11.0	593	2	US-09-252-991A-20441	Sequence 20441, A
18	124	10.9	264	2	US-09-107-433-2798	Sequence 2798, Ap
19	123	10.8	248	2	US-09-252-991A-22391	Sequence 22391, A
20	123	10.8	723	2	US-10-104-047-2246	Sequence 2246, Ap
21	123	10.8	1706	1	US-08-455-568-2	Sequence 2, Appli
22	123	10.8	1706	1	US-08-399-411-2	Sequence 2, Appli
23	123	10.8	1706	2	US-08-516-859A-2	Sequence 2, Appli
24	123	10.8	1706	2	US-09-586-472-2	Sequence 2, Appli
25	123	10.8	1706	2	US-09-528-706-2	Sequence 2, Appli
26	123	10.8	1706	2	US-10-024-450-2	Sequence 2, Appli
27	122	10.8	957	2	US-09-252-991A-20408	Sequence 20408, A
28	121.5	10.7	250	2	US-09-252-991A-27781	Sequence 27781, A
29	121.5	10.7	675	2	US-09-332-063-2	Sequence 2, Appli
30	121.5	10.7	675	2	US-09-332-063-3	Sequence 3, Appli
31	121	10.7	211	2	US-09-252-991A-26873	Sequence 26873, A
32	121	10.7	1485	2	US-09-762-569-4	Sequence 4, Appli
33	120.5	10.6	199	2	US-09-252-991A-26946	Sequence 26946, A
34	120.5	10.6	222	2	US-09-252-991A-32998	Sequence 32998, A
35	120.5	10.6	481	2	US-09-949-016-9748	Sequence 9748, Ap
36	120.5	10.6	855	2	US-09-949-016-7263	Sequence 7263, Ap
37	120.5	10.6	889	2	US-09-949-016-6036	Sequence 6036, Ap
38	120	10.6	442	2	US-08-834-306-52	Sequence 52, Appl
39	120	10.6	442	2	US-08-993-674A-52	Sequence 52, Appl
40	120	10.6	442	2	US-09-256-976-52	Sequence 52, Appl
41	120	10.6	605	2	US-09-949-016-8269	Sequence 8269, Ap
42	120	10.6	630	2	US-09-949-016-7270	Sequence 7270, Ap
43	120	10.6	802	2	US-09-823-240A-2	Sequence 2, Appli
44	119	10.5	316	2	US-09-252-991A-25345	Sequence 25345, A
45	118.5	10.4	294	2	US-09-252-991A-30882	Sequence 30882, A

ALIGNMENTS

```
RESULT 1
US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Query Match           13.4%;  Score 152; DB 2;  Length 2972;
Best Local Similarity 25.7%;  Pred. No. 0.00024;
Matches   69; Conservative  27; Mismatches  92; Indels  27; Gaps  12;

Qy     2 PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS 58
           |:||:|   |::::|       :|:::
Db  1189 PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVVPAAPGPPSLQPSGASPSAS 1244

Qy    59 A-AAGIA------                 -----SSAVEPVCGDA-APACLLRTPLRGLL 89
           :                              ::|
Db  1245 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPPPSAP 1304

Qy    90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP---- 144
            ::                         :::::              :|:::
Db  1305 NPAPAQASLLAP--------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1351

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT------ACPAPLPVVLVAPRST------ 182
              :|:                  ::          ::|
Db  1352 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQMLPAPVPS 1411

Qy   183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
            ::                |
Db  1412 PLPSPASTQTL---ALAPALAPTLGGSS 1436

RESULT 2
US-09-579-181-1
; Sequence 1, Application US/09579181
; Patent No. 6365372
```

```
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3118
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-1

Query Match       13.4%;  Score 152;  DB 2;  Length 3118;
Best Local Similarity   25.7%;  Pred. No. 0.00025;
Matches  69;  Conservative  27;  Mismatches  92;  Indels  80;  Gaps  12;

Qy     2 PHOAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS 58
           |:|:|     |  | :||:|:||                  |||||
Db  1335 PAPASAPLTIPISAPLTVSASG---PALLTSVTPPLAPVVPAAPGPPSLQPSGASPSAS 1390

Qy    59 A-AAGIA------------SSAVEPVCGDA-APACLLRTPLRGLL 89
           | |:|                : :||         |:
Db  1391 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1450

Qy    90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGPP----- 144
            : |       :                              :| :
Db  1451 NPAPAQASLLAP---------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1497

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT------ACPAPLPVVLVAPRST------- 182
               |                            ||:|| |:
Db  1498 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTVLAPSSTQTMLPAPVPS 1557

Qy   183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
            |||                |||:||
Db  1558 PLPSPASTQTL---ALAPALAPTLGGSS 1582

RESULT 3
US-09-252-991A-18097
; Sequence 18097, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18097
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18097

Query Match       11.9%;  Score 134.5;  DB 2;  Length 308;
Best Local Similarity   29.0%;  Pred. No. 0.00059;
Matches  53;  Conservative  18;  Mismatches  83;  Indels  29;  Gaps  7;

Qy    38 PPAACSGDPGCGSGAGLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRST 97
          ||  |                :|  | :||                   | :||| 
Db   112 PPACPAGSSSLSSLAAADAAGGASGTASGRPGVVAGSSARAS--GRPLRTRPSRRSTRPS 169

Qy    98 MECPPALIVHP-PAGG--MASGSSQPWAAAASATPMLSSKASLCIPTRGPPPQPLMRTPA 153
```

```
Db   170 ISSRPAPSAQPASPAGGVPIRSNRSQPW--------:|:||||:-----SNASICSDCRS------RRPS 212
             :||||                   |:| |||||
Qy   154 ARSHWPIPHPCDTACPAPLPVVLVAPRSTILS-MSRTWTC--RRWAVAPCRAEKLMC 207
             ::||:                 |:|: ||:|| |||:|
Db   213 CASTRQSSMTCATRTPASSPTIRAAPFSECAARMQASSWSAWVGLRSSAIRPALSTWVWA 272

Qy   208 SSS 210
             |:|
Db   273 SAS 275

RESULT 4
US-09-949-016-11382
; Sequence 11382, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11382
; LENGTH: 1540
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11382

Query Match       11.9%;  Score 134.5;  DB 2;  Length 1540;
Best Local Similarity   28.3%;  Pred. No. 0.0037;
Matches  54;  Conservative  21;  Mismatches  71;  Indels  45;  Gaps  9;

Qy    19 MGQRALPSSLALLSRPLSPP-----PAACSG-----DPCGCSGAGLPSASAAAGIASSA 67
           : |                      |||||         | | ||||||
Db   745 MLQKVLLNEYNGIDLPVENPADGTRSPSPCKSLEAQPDDLGPGSGFPAPTV-----ES 798

Qy    68 VEPVCGDAAPACLLRTP------LRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW 121
            : ||:|  ||             |||| ||               | ||      |
Db   799 TPDVC-PSSPA--LQTPSLSSGQLPPLLIPTDPSSPPPCCPPVLTVATPPPPLLPTVPLPA 855

Qy   122 AAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVV-----L 176
            :::| :           |         :     | :|:|:             |
Db   856 PSSSASP------HPCPSPLSNA-TAQSLPILSPTVSPSPSPIPPVEPLMSA 901

Qy   177 VAPRSTILSMS 187
               :||
Db   902 ASPGPPTLSSS 912

RESULT 5
US-09-949-016-11383
; Sequence 11383, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11383
; LENGTH: 1540
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11383

Query Match        11.9%;   Score 134.5; DB 2; Length 1540;
Best Local Similarity 28.3%; Pred. No. 0.0037;
Matches   54; Conservative 21; Mismatches 71; Indels 45; Gaps 9;

Qy   19 MGQRALPSSLALLSRPLSPP------PAACSG-----DPGCGSGAGLPSASAAAGIASSA 67
         |:|::|| :|:: |:|:  :    ||::  :   :||| ||  |||: :
Db  745 MLQKVLLNEYNGIDLPVENPADGTRSPSPCKSLEAQPDDLGPGSGFPAPTV-----ES 798

Qy   68 VEPVCGDAAPACLLRTP------LRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW 121
         |:|:|| :| |||:    ::|||| |:|
Db  799 TPDVC-PSSPA--LQTPSLSSGQLPPLLIPTDPSSPPPCPPVLTVATPPPLLPTVPLPA 855

Qy  122 AAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVV-----L 176
         ::|||      |:|:        ||:|:|
Db  856 PSSSASP---------------HPCPSPLSNA-TAQSPLPILSPTVSPSPIPPVEPLMSA 901

Qy  177 VAPRSTILSMS 187
         :| ||||
Db  902 ASPGPPTLSSS 912

RESULT 6
US-08-459-568-4
; Sequence 4, Application US/08459568
; Patent No. 5811304
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,568
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-459-568-4
```

```
Query Match        11.9%;   Score 134.5; DB 1; Length 1719;
Best Local Similarity 28.3%; Pred. No. 0.0041;
Matches   54; Conservative 21; Mismatches 71; Indels 45; Gaps 9;

Qy   19 MGQRALPSSLALLSRPLSPP------PAACSG-----DPGCGSGAGLPSASAAAGIASSA 67
         |:|::|| :|:: |:|:  :    ||::  :   :||| ||  |||: :
Db  888 MLQKVLLNEYNGIDLPVENPADGTRSPSPCKSLEAQPDDLGPGSGFPAPTV-----ES 941

Qy   68 VEPVCGDAAPACLLRTP------LRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW 121
         |:|:|| :| |||:    ::|||| |:|
Db  942 TPDVC-PSSPA--LQTPSLSSGQLPPLLIPTDPSSPPPCPPVLTVATPPPLLPTVPLPA 998

Qy  122 AAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVV-----L 176
         ::|||      |:|:        ||:|:|
Db  999 PSSSASP---------------HPCPSPLSNA-TAQSPLPILSPTVSPSPIPPVEPLMSA 1044

Qy  177 VAPRSTILSMS 187
         :| ||||
Db 1045 ASPGPPTLSSS 1055

RESULT 7
US-08-399-411-4
; Sequence 4, Application US/08399411
; Patent No. 5831008
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,411
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-399-411-4

Query Match        11.9%;   Score 134.5; DB 1; Length 1719;
Best Local Similarity 28.3%; Pred. No. 0.0041;
Matches   54; Conservative 21; Mismatches 71; Indels 45; Gaps 9;

Qy   19 MGQRALPSSLALLSRPLSPP------PAACSG-----DPGCGSGAGLPSASAAAGIASSA 67
         |:|::|| :|:: |:|:  :    ||::  :   :||| ||  |||: :
Db  888 MLQKVLLNEYNGIDLPVENPADGTRSPSPCKSLEAQPDDLGPGSGFPAPTV-----ES 941

Qy   68 VEPVCGDAAPACLLRTP------LRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW 121
         |:|:|| :| |||:    ::|||| |:|
Db  942 TPDVC-PSSPA--LQTPSLSSGQLPPLLIPTDPSSPPPCPPVLTVATPPPLLPTVPLPA 998
```

```
Qy    122  AAASATPMLSSKASLCIPTRGPPQPLMRTPAARSHWPIPHPCDTACPAPLPVV-----L  176
                :::||:||:|                            |||||:|:||:||:|
Db    999  PSSSASP---------------------HPCPSPLSNA-TAQSPLPILSPTVSPSPIPPVEPLMSA  1044

Qy    177  VAPRSTILSMS  187
                :|:||
Db   1045  ASPGPPTLSSS  1055
```

RESULT 8
US-08-516-859A-4
; Sequence 4, Application US/08516859A
; Patent No. 6069231
; GENERAL INFORMATION:
;   APPLICANT: Huang, Shi
;   TITLE OF INVENTION: Retinoblastoma Protein - Interacting
;   TITLE OF INVENTION: Zinc Finger Proteins
;   NUMBER OF SEQUENCES: 106
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Campbell & Flores LLP
;   STREET: 4370 La Jolla Village Drive, Suite 700
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92122
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/516,859A
;   FILING DATE: 18-AUG-1995
;   CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/399,411
;   FILING DATE: 06-MAR-1995
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/292,683
;   FILING DATE: 18-AUG-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Campbell, Cathryn A.
;   REGISTRATION NUMBER: 31,815
;   REFERENCE/DOCKET NUMBER: P-LJ 1776
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619) 535-9001
;   TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1719 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-516-859A-4

Query Match   11.9%;   Score 134.5; DB 2; Length 1719;
Best Local Similarity 28.3%;   Pred. No. 0.0041;
Matches   54; Conservative 21; Mismatches 71; Indels 45; Gaps   9;

```
Qy     19  MGQRALPSSLALLSRPLSPP------PAACSG-----DPGCGSGAGLPSASAAAGIASSA  67
                : :|:|: |  |   :||                     ||:  | |:| :|
Db    888  MLQKVLLNEYNGIDLPVENPADGTRSPSPCKSLEAQPDPDLGPGSGFPAPTV------ES  941

Qy     68  VEPVCGDAAPACLLRTP------LRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW  121
                : :| || |  ||                ||::|                    :
Db    942  TPDVC-PSSPA--LQTPSLSSGQLPPLLIPTDPSSPPPCPPVLTVATPPPLLPTVPLPA  998

Qy    122  AAASATPMLSSKASLCIPTRGPPQPLMRTPAARSHWPIPHPCDTACPAPLPVV-----L  176
                :::||:||:|                            |||||:|:||:||:|
Db    999  PSSSASP---------------------HPCPSPLSNA-TAQSPLPILSPTVSPSPIPPVEPLMSA  1044

Qy    177  VAPRSTILSMS  187
                :|:||
Db   1045  ASPGPPTLSSS  1055
```

RESULT 9
US-09-586-472-4
; Sequence 4, Application US/09586472
; Patent No. 6323335
; GENERAL INFORMATION:
;   APPLICANT: Huang, Shi
;   TITLE OF INVENTION: Retinoblastoma Protein - Interacting
;   TITLE OF INVENTION: Zinc Finger Proteins
;   NUMBER OF SEQUENCES: 106
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Campbell & Flores LLP
;   STREET: 4370 La Jolla Village Drive, Suite 700
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92122
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/586,472
;   FILING DATE: 01-Jun-2000
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 09/528,706
;   FILING DATE: 17-MAR-2000
;   APPLICATION NUMBER: US 08/516,859
;   FILING DATE: 18-AUG-1995
;   APPLICATION NUMBER: US 08/399,411
;   FILING DATE: 06-MAR-1995
;   APPLICATION NUMBER: US 08/292,683
;   FILING DATE: 18-AUG-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Campbell, Cathryn A.
;   REGISTRATION NUMBER: 31,815
;   REFERENCE/DOCKET NUMBER: P-LJ 4130
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619) 535-9001
;   TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1719 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-586-472-4

Query Match   11.9%;   Score 134.5; DB 2; Length 1719;
Best Local Similarity 28.3%;   Pred. No. 0.0041;
Matches   54; Conservative 21; Mismatches 71; Indels 45; Gaps   9;

```
Qy     19  MGQRALPSSLALLSRPLSPP------PAACSG-----DPGCGSGAGLPSASAAAGIASSA  67
                : :|:|: |  |   :||                     ||:  | |:| :|
Db    888  MLQKVLLNEYNGIDLPVENPADGTRSPSPCKSLEAQPDPDLGPGSGFPAPTV------ES  941

Qy     68  VEPVCGDAAPACLLRTP------LRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW  121
                : :| || |  ||                ||::|                    :
Db    942  TPDVC-PSSPA--LQTPSLSSGQLPPLLIPTDPSSPPPCPPVLTVATPPPLLPTVPLPA  998

Qy    122  AAASATPMLSSKASLCIPTRGPPQPLMRTPAARSHWPIPHPCDTACPAPLPVV-----L  176
                :::||:||:|                            |||||:|:||:||:|
Db    999  PSSSASP---------------------HPCPSPLSNA-TAQSPLPILSPTVSPSPIPPVEPLMSA  1044

Qy    177  VAPRSTILSMS  187
                :|:||
Db   1045  ASPGPPTLSSS  1055
```

RESULT 10
US-09-528-706-4
; Sequence 4, Application US/09528706
; Patent No. 6468985
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/528,706
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/516,859
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-528-706-4

```
Query Match             11.9%;  Score 134.5;  DB 2;  Length 1719;
Best Local Similarity   28.3%;  Pred. No. 0.0041;
  Matches  54;  Conservative  21;  Mismatches  71;  Indels  45;  Gaps   9;

Qy   19 MGQRALPSSLALLSRPLSPP------PAACSG-----DPGCGSGAGLPSASAAAGIASSA 67
Db  888 MLQKVLLNEYNGIDLPVENPADGTRSPSPCKSLEAQPDPDLGPGSGFPAPTV------ES 941

Qy   68 VEPVCGDAAPACLLRTP------LRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW 121
Db  942 TPDVC-PSSPA--LQTPSLSSGQLPPLLIPTDPSSPPPCCPVLTVATPPPLLPTVPLPA 998

Qy  122 AAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVV-----L 176
Db  999 PSSSASP------------------HPCPSPLSNA-TAQSPLPILSPTVSPSPIPPVEPLMSA 1044

Qy  177 VAPRSTILSMS 187
Db 1045 ASPGPPTLSSS 1055
```

RESULT 11
US-10-024-450-4
; Sequence 4, Application US/10024450
; Patent No. 6927030
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; APPLICANT: Chadwick, Robert B.
; TITLE OF INVENTION: Methods of Detecting and Treating
; TITLE OF INVENTION: Microsatellite-instability Positive Tumors Using RIZ
; FILE REFERENCE: P-LJ 5101
; CURRENT APPLICATION NUMBER: US/10/024,450
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/256,582
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1719
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-024-450-4

```
Query Match             11.9%;  Score 134.5;  DB 2;  Length 1719;
Best Local Similarity   28.3%;  Pred. No. 0.0041;
  Matches  54;  Conservative  21;  Mismatches  71;  Indels  45;  Gaps   9;

Qy   19 MGQRALPSSLALLSRPLSPP------PAACSG-----DPGCGSGAGLPSASAAAGIASSA 67
Db  888 MLQKVLLNEYNGIDLPVENPADGTRSPSPCKSLEAQPDPDLGPGSGFPAPTV------ES 941

Qy   68 VEPVCGDAAPACLLRTP------LRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW 121
Db  942 TPDVC-PSSPA--LQTPSLSSGQLPPLLIPTDPSSPPPCCPVLTVATPPPLLPTVPLPA 998

Qy  122 AAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVV-----L 176
Db  999 PSSSASP------------------HPCPSPLSNA-TAQSPLPILSPTVSPSPIPPVEPLMSA 1044

Qy  177 VAPRSTILSMS 187
Db 1045 ASPGPPTLSSS 1055
```

RESULT 12
US-09-252-991A-28087
; Sequence 28087, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28087
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28087

```
Query Match             11.8%;  Score 133.5;  DB 2;  Length 339;
Best Local Similarity   27.6%;  Pred. No. 0.0008;
  Matches  66;  Conservative  14;  Mismatches  92;  Indels  67;  Gaps  12;

Qy    4 QAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPA------ACSGDPGCGSGA----- 52
Db  102 RGAGPEDRHPRRRREARRLGLARS-RLSSFRRAPDPARAGLQLSAAGRTAAGSGADPPMA 160

Qy   53 ---GLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMEC--PPALIVH 107
Db  161 TRRGTPSWMSACANSSSASPGTC-----ATRLAMRRRTSKNVACGSTASCSTTPSKASL 215
```

```
Qy    108  PP----AGGMASGSSQPWAAAS-----ATPMLS------SKASLC------IPTRGPPPQP  147
               | ||||:|||||          :::        ||||||          || |||||||
Db    216  PAASRACAGCLASSSGMAWCASSIARREAIPPCCSPVSPGSSSTTCGSGASIGRRGKPSWP  275

Qy    148  LMRTPAARSHWPIP--------HPCDTACPAPLPVVLVAPRSTILSMSRTWTCRRWAVA   198
               ||||||                |||:
Db    276  TSNGAKRASTWPSTIRRIGVKATCSTGC-------------RCCRAWRC-RWAIA      317


RESULT 13
US-10-104-047-3402
; Sequence 3402, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3402
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3402

Query Match        11.3%;   Score 128;  DB 2;  Length 710;
Best Local Similarity   27.5%;  Pred. No. 0.0055;
Matches   64;  Conservative   22;  Mismatches  73;  Indels  74;  Gaps  12;

Qy      4  QAAAPVDQTPRTLATMGQRALPSSLALLSRP-----------------LSPPPAACSG    44
            :||||   ||||:||       |    | |:                  ||    |||
Db     96  RAPAPVPQ-PFSLPBPSQPILPSVLSLLGLPTPGPSHSDGSFNLLGSDAHLPPPTLSSG   154

Qy     45  D--------PGCGSGAGLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPT    92
                     | ::    ::| :|||||:||
Db    155  SPPQPRHPIQPSLPGTTSGS-LSSVPGAPAPPAASKAPV----VPSPVLQSPSEGLGMGA   209

Qy     93  GPRSTMECPPALIVHPPAGGMA----------SGSSQPWAAASATPM------------  129
                                            :||
Db    210  GP-----ACP----LPPLAGGEAFPFPSPBQGLALSGAGFP-GMLGALPLPLSLGQPPPSP   260

Qy    130  LSSKASLCIPTRG---PPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAP   179
            |||:|| |  ||     ||||||||||||||        |||:
Db    261  LLNHSLFGVLTGGGQPPPEPLLPPPGGPG--PPLAPGEPEGPSLLVASLLPP   311


RESULT 14
US-10-104-047-2232
; Sequence 2232, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2232
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2232

Query Match        11.2%;   Score 127;  DB 2;  Length 246;
Best Local Similarity   24.9%;  Pred. No. 0.002;
Matches   68;  Conservative   17;  Mismatches  60;  Indels  128;  Gaps  14;
```

```
Qy      7  APVDQTPRTLAT----MGQR-ALPSSLA-LLSRPLSPPPAACSGDPGCGSGAGLPSASA   59
            |||:|||||||    ||||  ||||||  :|||||||||| ||||   | |::  ||
Db      5  ASVDCTPPSLQTELLPPWGPRSSLPADCAPCLGTALLPPWGPCSSLPPCGPRSSLP----   60

Qy     60  AAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTM---ECPPAL--------I   105
            |||| |::  ||||                 :               ||||
Db     61  -----------GDRAPPS-LRTIALG--PPWGRPSLPADRAPPSLGTSLPPPSLRTV   103

Qy    106  VHPPAGGMASGSSQPWAA-----ASATPMLSSKASLC-----  137
            :||:                   ||:
Db    104  LLPPCG---LCFSLPWSSLDLPELCQVGPGRPFPGPAVALSFLTCFHSCCPRWALGFVLL   160

Qy    138  IPTR--------------GPPPQPLMRTPAARSHWPIPH   162
                              ||:||
Db    161  LKPFVLNLRGSRDWGIPSKELWLGAQKEGHRVLVGCPGSGSGPPPLPASISVQGSPG    220

Qy    163  PCDTACPAPLPVVLVAPRSTILSMSRTWTCRRW   195
            ||:
Db    221  PC----------SPLLSVS------AERW      233


RESULT 15
US-10-104-047-2243
; Sequence 2243, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2243
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2243

Query Match        11.2%;   Score 127;  DB 2;  Length 246;
Best Local Similarity   24.9%;  Pred. No. 0.002;
Matches   68;  Conservative   17;  Mismatches  60;  Indels  128;  Gaps  14;

Qy      7  APVDQTPRTLAT----MGQR-ALPSSLA-LLSRPLSPPPAACSGDPGCGSGAGLPSASA   59
            |||:|||||||    ||||  ||||||  :|||||||||| ||||   | |::  ||
Db      5  ASVDCTPPSLQTELLPPWGPRSSLPADCAPCLGTALLPPWGPCSSLPPCGPRSSLP----   60

Qy     60  AAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTM---ECPPAL--------I   105
            |||| |::  ||||                 :               ||||
Db     61  -----------GDRAPPS-LRTIALG--PPWGRPSLPADRAPPSLGTSLPPPSLRTV   103

Qy    106  VHPPAGGMASGSSQPWAA-----ASATPMLSSKASLC-----  137
            :||:                   ||:
Db    104  LLPPCG---LCFSLPWSSLDLPELCQVGPGRPFPGPAVALSFLTCFHSCCPRWALGFVLL   160

Qy    138  IPTR--------------GPPPQPLMRTPAARSHWPIPH   162
                              ||:||
Db    161  LKPFVLNLRGSRDWGIPSKELWLGAQKEGHRVLVGCPGSGSGPPPLPASISVQGSPG    220

Qy    163  PCDTACPAPLPVVLVAPRSTILSMSRTWTCRRW   195
            ||:
Db    221  PC----------SPLLSVS------AERW      233


Search completed: March 16, 2006, 15:58:03
Job time : 49 secs
```

```
          GenCore version 5.1.7
       Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on:    March 16, 2006, 15:57:25 ; Search time 165 Seconds
                               (without alignments)
                     536.847 Million cell updates/sec

Title:       US-09-989-890-238
Perfect score: 1134
Sequence:    1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:    1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters:    1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Published Applications AA Main:*
              1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
              2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
              3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
              4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
              5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
              6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	1134	100.0	212	3	US-09-989-890-238	Sequence 238, App
2	1069.5	94.3	213	3	US-09-989-920-206	Sequence 206, App
3	1069.5	94.3	213	3	US-09-989-920-251	Sequence 251, App
4	1065.5	94.0	218	4	US-10-074-475-261	Sequence 261, App
5	161.5	14.2	556	4	US-10-437-963-137918	Sequence 137918,
6	157	13.8	487	4	US-10-437-963-137896	Sequence 137896,
7	154	13.6	1480	4	US-10-408-765A-2239	Sequence 2239, Ap
8	152	13.4	2971	4	US-10-146-473-50	Sequence 50, Appl
9	152	13.4	2971	5	US-10-732-923-8849	Sequence 8849, Ap
10	149.5	13.2	1828	4	US-10-221-625-15	Sequence 15, Appl
11	147.5	13.0	437	4	US-10-437-963-197867	Sequence 197867,
12	144.5	12.7	454	4	US-10-437-963-102727	Sequence 102727,
13	143.5	12.7	373	4	US-10-437-963-149274	Sequence 149274,
14	143	12.6	166	4	US-10-437-963-156672	Sequence 156672,
15	142.5	12.6	180	4	US-10-437-963-190430	Sequence 190430,
16	141.5	12.5	625	4	US-10-437-963-143626	Sequence 143626, Ap
17	141	12.4	1008	4	US-10-276-774-1897	Sequence 1897, Ap
18	139.5	12.3	569	4	US-10-437-963-149276	Sequence 149276,
19	138	12.2	437	4	US-10-437-963-168762	Sequence 168762,
20	138	12.2	883	4	US-10-437-963-116012	Sequence 116012,
21	136.5	12.0	415	5	US-10-450-763-60509	Sequence 60509, A
22	136	12.0	309	6	US-10-425-115-198647	Sequence 198647,
23	135.5	11.9	842	6	US-11-097-143-26685	Sequence 26685, A
24	135.5	11.9	864	6	US-11-097-143-40749	Sequence 40749, A
25	135	11.9	693	4	US-10-108-260A-3849	Sequence 3849, Ap
26	135	11.9	3164	4	US-10-669-161-72	Sequence 72, Appl
27	134.5	11.9	161	4	US-10-767-701-34009	Sequence 34009, A
28	134.5	11.9	261	4	US-10-425-115-262328	Sequence 262328,
29	134.5	11.9	328	4	US-10-425-115-198645	Sequence 198645,
30	134.5	11.9	983	3	US-09-864-761-38000	Sequence 38000, A
31	134.5	11.9	1518	4	US-10-142-650-2	Sequence 2, Appli
32	134.5	11.9	1719	4	US-10-024-450-4	Sequence 4, Appli
33	134.5	11.9	1719	4	US-10-142-650-1	Sequence 1, Appli
34	134.5	11.9	1719	6	US-11-077-465-4	Sequence 4, Appli
35	134	11.8	215	4	US-10-437-963-162087	Sequence 162087,
36	134	11.8	215	4	US-10-425-115-315380	Sequence 315380,
37	133.5	11.8	507	4	US-10-437-963-124879	Sequence 124879,
38	133	11.7	231	4	US-10-767-701-38778	Sequence 38778, A
39	133	11.7	1013	5	US-10-450-763-38471	Sequence 38471, A
40	132.5	11.7	369	4	US-10-437-963-162013	Sequence 162013,
41	132.5	11.7	554	4	US-10-437-963-119377	Sequence 119377,
42	132.5	11.7	818	4	US-10-369-493-18031	Sequence 18031, A
43	132	11.6	230	4	US-10-425-115-296496	Sequence 296496,
44	132	11.6	269	4	US-10-425-114-38093	Sequence 38093, A
45	132	11.6	353	4	US-10-437-963-203523	Sequence 203523,

```
                              ALIGNMENTS

RESULT 1
US-09-989-890-238
; Sequence 238, Application US/09989890
; Publication No. US20040166105A1
; GENERAL INFORMATION:
;  APPLICANT: Salceda, Susana
;  APPLICANT: Macina, Roberto
;  APPLICANT: Recipon, Herve
;  APPLICANT: Pluta, Jason
;  APPLICANT: Sun, Yongming
;  APPLICANT: Liu, Chenghua
;  TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
;  FILE REFERENCE: DEX-0287
;  CURRENT APPLICATION NUMBER: US/09/989,890
;  CURRENT FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: 60/252,509
;  PRIOR FILING DATE: 2000-11-22
;  NUMBER OF SEQ ID NOS: 280
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 238
;  LENGTH: 212
;  TYPE: PRT
;  ORGANISM: Homo sapien
US-09-989-890-238

Query Match          100.0%;  Score 1134;  DB 3;  Length 212;
Best Local Similarity 100.0%;  Pred. No. 6.7e-73;
Matches 212;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60

Qy  61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQP 120

Qy 121 WAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 WAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPR 180

Qy 181 STILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 212
         ||||||||||||||||||||||||||||||||
Db 181 STILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 212

RESULT 2
US-09-989-920-206
; Sequence 206, Application US/09989920
; Patent No. US20020172957A1
```

```
; GENERAL INFORMATION:
;   APPLICANT: Macina, Roberto
;   APPLICANT: Recipon, Herve
;   APPLICANT: Chen, Sei-Yu
;   APPLICANT: Sun, Yongming
;   APPLICANT: Liu, Chenghua
;   TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
;   FILE REFERENCE: DEX-0291
;   CURRENT APPLICATION NUMBER: US/09/989,920
;   CURRENT FILING DATE: 2001-11-21
;   PRIOR APPLICATION NUMBER: 60/252,500
;   PRIOR FILING DATE: 2000-11-22
;   NUMBER OF SEQ ID NOS: 284
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 206
;   LENGTH: 213
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-989-920-206

Query Match             94.3%;  Score 1069.5;  DB 3;  Length 213;
Best Local Similarity   95.3%;  Pred. No. 2.5e-68;
Matches   203;  Conservative    1;  Mismatches    8;  Indels    1;  Gaps    1;

Qy    1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60

Qy   61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMA-SGSSQ 119
        |||||||||| |||||||||||||||||||||||||||||||||||||||  :||
Db   61 AGIASSAVEAVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPTGGMARRAASQ 120

Qy  120 PWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAP 179
        |||||||||||||||||||||   |||||||||||||||||||  |||||||||||||||
Db  121 PWAAASATPMLSSKASLCIPTERPPPQPLMRTPAARSHWPIPHPASTACPAPLPVVLVAP 180

Qy  180 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 212
        |||||||||||||||||||||||||||||||||
Db  181 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 213


RESULT 3
US-09-989-920-251
; Sequence 251, Application US/09989920
; Patent No. US2002017295A1
; GENERAL INFORMATION:
;   APPLICANT: Macina, Roberto
;   APPLICANT: Recipon, Herve
;   APPLICANT: Chen, Sei-Yu
;   APPLICANT: Sun, Yongming
;   APPLICANT: Liu, Chenghua
;   TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
;   FILE REFERENCE: DEX-0291
;   CURRENT APPLICATION NUMBER: US/09/989,920
;   CURRENT FILING DATE: 2001-11-21
;   PRIOR APPLICATION NUMBER: 60/252,500
;   PRIOR FILING DATE: 2000-11-22
;   NUMBER OF SEQ ID NOS: 284
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 251
;   LENGTH: 213
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-989-920-251

Query Match             94.3%;  Score 1069.5;  DB 3;  Length 213;
Best Local Similarity   95.3%;  Pred. No. 2.5e-68;
Matches   203;  Conservative    1;  Mismatches    8;  Indels    1;  Gaps    1;

Qy    1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60

Qy   61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMA-SGSSQ 119
        |||||||||| |||||||||||||||||||||||||||||||||||||||  :||
Db   61 AGIASSAVEAVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPTGGMARRAASQ 120

Qy  120 PWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAP 179
        |||||||||||||||||||||   |||||||||||||||||||  |||||||||||||||
Db  121 PWAAASATPMLSSKASLCIPTERPPPQPLMRTPAARSHWPIPHPASTACPAPLPVVLVAP 180

Qy  180 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 212
        |||||||||||||||||||||||||||||||||
Db  181 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 213


RESULT 4
US-10-074-475-261
; Sequence 261, Application US/10074475
; Publication No. US2003009289BA1
; GENERAL INFORMATION:
;   APPLICANT: Salceda, Susana
;   APPLICANT: Macina, Roberto
;   APPLICANT: Hu, Ping
;   APPLICANT: Recipon, Herve
;   APPLICANT: Karra, Kalpana
;   APPLICANT: Cafferkey, Robert
;   APPLICANT: Sun, Yongming
;   APPLICANT: Liu, Chenghua
;   TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
;   TITLE OF INVENTION: Genes and Proteins
;   FILE REFERENCE: DEX-0313
;   CURRENT APPLICATION NUMBER: US/10/074,475
;   CURRENT FILING DATE: 2002-02-13
;   PRIOR APPLICATION NUMBER: 60/268,292
;   PRIOR FILING DATE: 2001-02-13
;   NUMBER OF SEQ ID NOS: 295
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 261
;   LENGTH: 218
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-074-475-261

Query Match             94.0%;  Score 1065.5;  DB 4;  Length 218;
Best Local Similarity   95.3%;  Pred. No. 5e-68;
Matches   203;  Conservative    1;  Mismatches    8;  Indels    1;  Gaps    1;

Qy    1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAA 65

Qy   61 AGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMA-SGSSQ 119
        |||||||||| |||||||||||||||||||||||||||||||||||||||  :||
Db   66 AGIASSAVEAVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVQPPAGGMARRAASQ 125

Qy  120 PWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAP 179
        |||||||||||||||||||||   |||||||||||||||||||  |||||||||||||||
Db  126 PWAAASATPMLSSKASLCIPTERPPPQPLMRTPAARSHWPIPHPASTACPAPLPVVLVAP 185

Qy  180 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 212
        |||||||||||||||||||||||||||||||||
Db  186 RSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS 218


RESULT 5
US-10-437-963-137918
; Sequence 137918, Application US/10437963
; Publication No. US2004012334A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
```

```
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
;   SEQ ID NO 137918
;   LENGTH: 556
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(556)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_39355C.1.pep
US-10-437-963-137918

  Query Match       14.2%;   Score 161.5;   DB 4;   Length 556;
  Best Local Similarity   25.8%;   Pred. No. 0.0019;
  Matches   71;   Conservative   15;   Mismatches 104;   Indels   85;   Gaps    8;

Qy     1  SPHQAAAPVDQTPRTLATMGQRAL----------------PSSLALLSRPLS--PPPA   40
             |||: :|    | | |||: ::| |                 |||||||||     |||
Db   210  SVHPAXAPEPRAPETLAPAPACALPTVNPTTAMVPRGGRPPASASLSRPLSWAAPPA  269

Qy    41  ACSGDPGCGS----------------GAGLPSASAAAGIASSAVEPVCGDAAPA----   78
               | |   :                 |  |::  |  |||||||||||||| :
Db   270  PAPXPLGPAQPHXTPRAAPKPPWPPPRTAALPSSSPRSPCPLSSRRPPRRGAAPTXPTT  329

Qy    79  -----CLLRTPLRGLLKPTGPRSTMECPPAL----------IVHPPAGGMASGSSQ  119
                |  |   |   |  ||  |                    ||||||||||||||: :
Db   330  KPPPPPCQPPPPTTPAPTPTTPPTPWTPPPATPTPPTPWTPTPATPPTPATPATPTT  389

Qy   120  PWAAASATPMLSKSKASLCIP----------TRGPPPQPLMR---TPAARSHWPIPHPCDTA  167
           ||| ||||||:|| |  ||| |         |||| ||||:   ||||| :  :
Db   390  PPTPAPAPPPATPAARCCPASPTSTPRSASAPPSRPXARPQPRPPPRSPWPPPPRTAA  449

Qy   168  CPAPLPVVLVAPRSTILSMSRTWTCRRWAVAPCRA  202
           | |||||||||||||:||||:|  |   |
Db   450  LPSP----------RSSWRRWPCPRSRRPRRA  472

RESULT 6
US-10-437-963-137896
; Sequence 137896, Application US/10437963
; Publication No. US2004012343A1
GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
;   SEQ ID NO 137896
;   LENGTH: 487
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_39355C.1.pep
US-10-437-963-137896

  Query Match       13.8%;   Score 157;   DB 4;   Length 487;
```

```
  Best Local Similarity   29.0%;   Pred. No. 0.0034;
  Matches   78;   Conservative   20;   Mismatches   85;   Indels   86;   Gaps   15;

Qy     1  SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPL-SPPPAAC----SGDPGCGSGAGLP   55
           | :  : ||   :   || |   :  |  |      |    ||
Db   221  SPMSPSAPMPSTPST--TRAPRATRASTTL--PLPGPPPRAMRPAPAPAPAAAAAPMP  275

Qy    56  SASAAAGIASSAVEPVCGDAAPACLLRT-PLRGLLKPTGPRSTMECPPA----------LIV  106
           :: |||   :  :      | || :|  : :     :    :    : ::        |:
Db   276  RSTAAATSPLTWTSPLTSAATTSPLTWTSPRTSPRTSSPSSTATTPPAPTPTPRLTLAV  335

Qy   107  HPPAGGMASGS----SQWAAAS--------------ATPMLSSK---ASLCIP------  139
               | :: |       :| ||                ||  :    |:|  :
Db   336  AAAPATMTSTAPLPEPWPWLAASTPPPIPALRLTLAATPATRSRSTTAPAAVPGVTAAS  395

Qy   140  --TRGPP----------------------------PQPLMR----TPAARS-HWPIPHPCDTACPA---  170
            ||                              ||| :     ||| :  ||||||||||||
Db   396  SSTTTPPLLTTMPVALLTAPPSATRSIPLPLSRPTTATPATRSIPLPSPKPRPTATPAAR  455

Qy   171  ----PLPVVLVAP---------RSTILSMSR  188
              ||||:| |          ||::| :|
Db   456  STSTPLPLPAPAPPWLMTRARLLMTAR  484

RESULT 7
US-10-408-765A-2239
; Sequence 2239, Application US/10408765A
; Publication No. US20040101874A1
GENERAL INFORMATION:
;   APPLICANT: Ghosh, Soumitra S.
;   APPLICANT: Fahy, Eoin D.
;   APPLICANT: Zhang, Bing
;   APPLICANT: Gibson, Bradford W.
;   APPLICANT: Taylor, Steven W.
;   APPLICANT: Glenn, Gary M.
;   APPLICANT: Warnock, Dale E.
;   TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
;   TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
;   FILE REFERENCE: 660088.465
;   CURRENT APPLICATION NUMBER: US/10/408,765A
;   CURRENT FILING DATE: 2003-04-04
;   NUMBER OF SEQ ID NOS: 3077
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 2239
;   LENGTH: 1480
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-10-408-765A-2239

  Query Match       13.6%;   Score 154;   DB 4;   Length 1480;
  Best Local Similarity   25.7%;   Pred. No. 0.017;
  Matches   69;   Conservative   27;   Mismatches   92;   Indels   80;   Gaps   12;

Qy     2  PHQAAAPVD--QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS   58
           | ||:|   : ||   |  ::    |  ||::   |||     |   ||||||::|||
Db  1018  PAPASAPLTIPISAPLTVSASG----PALLTSVTPLAPVVPAAPGPPSLAPSGASPSAS  1073

Qy    59  A-AAGIA----------------SSAVEPVCGDA-APACLLRTPLRGLL   89
           | ||: |                ||| :  ||   ||| ||:|| | ||
Db  1074  ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPPSAP  1133

Qy    90  KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP------  144
            :|   ::    : |  :     |::|||:  |||: |   :::     |:|||
Db  1134  NPAPAQASLLAP---------ASSASQALATPLA-PWAAPQTAILAPSPAPPLAPLP  1180

Qy   145  ---PQPLMRTPAARSHWPIP------HPCDT--------ACPAPLPVVLVAPRST--------  182
              || |   |  |: :||         |  |         | ||||||:  |||
Db  1181  VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS  1240
```

Qy   183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
          : :|| | |||| |||| ||||
Db   1241 PLPSPASTQTL--ALAPALAPTLGGSS 1265

RESULT 8
US-10-146-473-50
; Sequence 50, Application US/10146473
; Publication No. US20030108888A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Gout, Ivan
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Gure, Ali
; APPLICANT: Chen, Yao-tseng
; APPLICANT: Old, Lloyd
; TITLE OF INVENTION: Breast Cancer Antigens
; FILE REFERENCE: L00461/70130(JRV)
; CURRENT APPLICATION NUMBER: US/10/146,473
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/291,150
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 50
; LENGTH: 2971
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-146-473-50

Query Match   13.4%; Score 152; DB 4; Length 2971;
Best Local Similarity   25.7%; Pred. No. 0.045;
Matches   69; Conservative 27; Mismatches 92; Indels 80; Gaps 12;

Qy   2 PHQAAAPVD---QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS 58
          |::|:| | :|:|: |:|::||||||
Db   1189 PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVVPAAPGPPSLQPSGASPSAS 1244

Qy   59 A-AAGIA------------------------SSAVEPVCGDA-APACLLRTPLRGLL 89
          | :|:| :|::
Db   1245 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1304

Qy   90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP----- 144
                           :::| || | |:| || :: :|
Db   1305 NPAPAQASLLAP-----------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1351

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT-------ACPAPLPVVLVAPRST-------- 182
             || : |:| |: || :|| |:|| |
Db   1352 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1411

Qy   183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
          : :|| | |||| |||| ||||
Db   1412 PLPSPASTQTL--ALAPALAPTLGGSS 1436

RESULT 9
US-10-732-923-8849
; Sequence 8849, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8849
; LENGTH: 2971
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-732-923-8849

Query Match   13.4%; Score 152; DB 5; Length 2971;
Best Local Similarity   25.7%; Pred. No. 0.045;
Matches   69; Conservative 27; Mismatches 92; Indels 80; Gaps 12;

Qy   2 PHQAAAPVD---QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS 58
          |::|:| | :|:|: |:|::||||||
Db   1189 PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVVPAAPGPPSLQPSGASPSAS 1244

Qy   59 A-AAGIA------------------------SSAVEPVCGDA-APACLLRTPLRGLL 89
          | :|:| :|::
Db   1245 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPFPSAP 1304

Qy   90 KPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPP----- 144
                           :::| || | |:| || :: :|
Db   1305 NPAPAQASLLAP-----------------ASSASQALATPLA-PMAAPQTAILAPSPAPPLAPLP 1351

Qy   145 ---PQPLMRTPAARSHWPIP--HPCDT-------ACPAPLPVVLVAPRST-------- 182
             || : |:| |: || :|| |:|| |
Db   1352 VLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLAPSSTQTMLPAPVPS 1411

Qy   183 -ILSMSRTWTCRRWAVAPCRAEKLMCSS 209
          : :|| | |||| |||| ||||
Db   1412 PLPSPASTQTL--ALAPALAPTLGGSS 1436

RESULT 10
US-10-221-625-15
; Sequence 15, Application US/10221625
; Publication No. US20040033942A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: BANDMAN, Olga
; APPLICANT: TANG, Y. Tom
; APPLICANT: MATHUR, Preete
; APPLICANT: SHAH, Purvi
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: REDDY, Roopa
; TITLE OF INVENTION: TRANSCRIPTION FACTORS
; FILE REFERENCE: PF-0761 PCT
; CURRENT APPLICATION NUMBER: US/10/221,625
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 1828
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US2004003942A1 879500CD1
US-10-221-625-15

Query Match   13.2%; Score 149.5; DB 4; Length 1828;
Best Local Similarity   26.6%; Pred. No. 0.042;
Matches   65; Conservative 22; Mismatches 88; Indels 69; Gaps 11;

Qy   2 PHQAAAPVD---QTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSAS 58
          |::|:| | :|:|: |:|::||||||
Db   84 PAPASAPLTIPISAPLTVSASG----PALLTSVTPPLAPVVPAAPGPPSLAPSGASPSAS 139

Qy   59 A-AAGIA------------------------SSAVEPVCGDA-APACLLRTPLRGLL 89
          | :|:| :|::
Db   140 ALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPACSPVLVPASALASPF----- 195

Qy   90 KPTGPRSTMECPPALIVHP-PAGGMASGSSQ---PWAAASATPMLSSKASLCIPTRGPP 145

```
Db    196 -PSAPNPAPPLAPLPVLAPSPGAAPVLASSQTPVPVMAPSSTP----GTSLASASPVPAP 250
              |: |       ::  |          |||       |||:|
Qy    146 QPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPRSTILsMSRTWTCRRWAVAPCRABKL 205
              |:|            |||:|  | |:|           |:||
Db    251 TPVL--------APSSTQTMLPAPVPSPLPSPAST--------QTLALAPALAPTL 290

Qy    206 MCSS 209
Db    291 GGSS 294
```

RESULT 11
US-10-437-963-197867
; Sequence 197867, Application US/10437963
; Publication No. US2004012334341
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 197867
;   LENGTH: 442
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(442)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_93584C.1.pep
US-10-437-963-197867

Query Match   13.0%;   Score 147.5; DB 4; Length 442;
Best Local Similarity   26.6%;   Pred. No. 0.015;
Matches   65;   Conservative   14;   Mismatches   104;   Indels   61;   Gaps   9;

```
Qy      2 PHQAAAPVDQTPR-TLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGS----------- 50
             ||:|  |       :|   |       ||          ||  ::|
Db     37 PLRAPPPPPPXPRPSSPPLPSRAPPPPLP-LXRPSPPPPSRAHPPPASSPXPSQPAPPP 95
             |                      |               ||  |  :
Qy     51 -----GAGLPSASAAAGIASSAVEP-----VCGDAAPACLL----------RTPLRGL 88
                ||                          |    |           |||:||
Db     96 RPSLTGVASPRPSPPAAPXSSPSPPPPIEYAAAAGCACFLSHIARCFHLPPPRAPRRG- 154
             |  |  |     |  |   |                   |   |
Qy     89 LKPTGPRSTMECPPALIVHPPAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGP---- 143
             ||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db    155 PSPPSPVVRALPAAALAPPP---IPSASQPRACHPRXPRSVPHPSRTLPPTAPXSKTP 211
                        ||||||           |     |  ||||  |     |
Qy    144 ---PPQPLMRT------------PAARSHWPIPHPCDTACPAPLPVVLVAPRSTILS 185
                  ||||||            | |:|  | |:|           |||::  |
Db    212 PHPNPPPPLTHTHTAAPQVAPPAHSPAASPPPPDPYPSSRALPCPYPLLFSSSYPPIFS 271

Qy    186 MSRT 189
             ::
Db    272 PAPT 275
```

RESULT 12
US-10-437-963-102727
; Sequence 102727, Application US/10437963
; Publication No. US2004012334341
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 102727
;   LENGTH: 454
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(454)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_100222C.1.pep
US-10-437-963-102727

Query Match   12.7%;   Score 144.5; DB 4; Length 454;
Best Local Similarity   28.4%;   Pred. No. 0.025;
Matches   57;   Conservative   16;   Mismatches   69;   Indels   59;   Gaps   9;

```
Qy     37 PPPAACSGDPG----------CGSGAGLPSASAAAGIASSAVEPVC-----GDAAPACLLR 82
             ||          |          ::|                  ||        |||
Db    145 PPPXPPPSDPGGPRPAAXPCSSPAPVPSPPASSSAPAAANPGCPPPASPPTAPAAVVE 204
                                           :  |  |
Qy     83 ----------TPLRGLLKPTGPRSTMECPPALIVHPPAGGMASG 116
                      |||  |  ||||||||||||||||||||||||
Db    205 GQRSPTPRAHSPDPPPLRAGHRRXPQPLPRAADPALPKAAPPNPPXPTPSPPPSPPPTR 264
             |      |   |   |         |    ||   |  :|
Qy    117 SSQPWAAASATPMLSSKASLCIPT--------RGPPP----QPLMR-TPAA---RSHWP 159
             |||||||||||||||||||||||            ||         ||  |:||
Db    265 SS-PTTSSPPTPVARSPMTPSVPXDLPAPPTRPRGPPPHRMASPLPRSTPPALIPASPPP 323
                   |     |     |     || |  |   |   |   |
Qy    160 IPHPCDTACPAPL-PVVLVAP 179
             || |:|  | :|||||||||||
Db    324 LPLPPXTATPVVVPPVPVAP 344
```

RESULT 13
US-10-437-963-149274
; Sequence 149274, Application US/10437963
; Publication No. US2004012334341
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 149274
;   LENGTH: 373
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(373)

```
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_49623C.1.pep
US-10-437-963-149274

Query Match         12.7%;     Score 143.5; DB 4;   Length 373;
Best Local Similarity   29.2%;     Pred. No. 0.024;
Matches    63;  Conservative   13;  Mismatches   89;  Indels   51;  Gaps   11;

Qy     1 SPHQAAAPVDQTPRTL-ATMGQRALPSSLALL--------SRPLSPPPAAC---SGDP 46
           |||   |||||  |   :|||                  :|:|     ||    ||||
Db    81 SPXPPPPXPAPPRPLPSSLPPPAPSTTTTTATPPPXPSAPLADPPAPLPPPPSSXP 140

Qy    47 GCCSGAGLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIV 106
           |||       :|:          :||          :|
Db   141 SSPSPPPAPSPPPPXPSSPSAPSP---PPAPSSPPXPP-----SPPHPRPTKQMPP---- 188

Qy   107 HPPAGGMAS--GSSQPWAAAAATPMLSSKASLCIPT------RGPPPQPLMR-------- 150
           ||       | :    ||           :|:|        :|
Db   189 -PPASSRRSTPRSXSPPWPASPSPPHSSSSPSAASSTARPIRRPPPPTPTVKPTSSTMVV 247

Qy   151 TPAAR-------SHWPIPHPCDTACPAPLPVVLVAP 179
           ||                 |     ||: | | |
Db   248 TPPARPSPSSRPTSTPPPPAGLP-PLPLPLPLPLP 282


RESULT 14
US-10-437-963-156672
;   Sequence 156672, Application US/10437963
;   Publication No. US20040123343A1
;   GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE:  2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
;   SEQ ID NO 156672
;   LENGTH: 166
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_56316C.1.pep
US-10-437-963-156672

Query Match         12.6%;     Score 143; DB 4;   Length 166;
Best Local Similarity   34.1%;     Pred. No. 0.012;
Matches    63;  Conservative   10;  Mismatches   72;  Indels   40;  Gaps   11;

Qy    32 SRPLSPPPAACSGDPGCGSGAGLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKP 91
           ||||||||:                    |      :|  :||||||||||||||
Db     7 SPPSPPPPRA----PPCTPTTGPPSE-----ASSPPRPLHADRRPT--IRSPLSAAPPP 54

Qy    92 TGPRSTMECPP--ALIVHPPAGGMASGSSQPWAAASATPMLSSKASLCIPTRGPPPQPLM 149
           ||  ||  |||  | |           ||      || |:    |:|
Db    55 SRSRSRLELPITRALRRSPP------TGSPPPLSACSA---SAARYRLVPLRHLPPGELC 105

Qy   150 R-------TPAARSHWPIPHPCDTACPAPLPVVLVAPRSTILSMSR---TWTCRRWAV 197
           |        ||          |  || :|:      ||:  |
Db   106 RRPPALPLPLPPAARGAPRP-PAARGAPPPP---VAP-STFVSAGRGRMRWRLTCCGS 160

Qy   198 APCRA 202
           |||||
Db   161 ATCAA 165


RESULT 15
US-10-437-963-190430
;   Sequence 190430, Application US/10437963
;   Publication No. US20040123343A1
;   GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE:  2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
;   SEQ ID NO 190430
;   LENGTH: 180
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(180)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_86846C.1.pep
US-10-437-963-190430

Query Match         12.6%;     Score 142.5; DB 4;   Length 180;
Best Local Similarity   29.1%;     Pred. No. 0.014;
Matches    59;  Conservative   17;  Mismatches   62;  Indels   65;  Gaps   10;

Qy     6 AAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGS-----GAGLPSASA 59
           |||   ||    |   |  :|::|| |      | :|||||          ::
Db    17 AAPPRRRPR------RAPTSMASPRRPRPAAPPPATAAGSQATSSSWRTPAGGGAPPSSP 69

Qy    60 AAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGP---RSTMECPPALIVHPPAGGMAS 115
           |||         || |||  |   |              ::|
Db    70 AATPSTS---PSAAPI--------------------PWAPSAIGRTTTPAPAA------- 102

Qy   116 GSSQPWAAASATPMLSSKASLCIPTRGPPPQP----LMRTPAARSHWPIPHPCDTACPAP 171
           |||  :||            ::||            ::
Db   103 ---PSSPAAATPSMS------------PSAATPYPQSAILPRAASSRDRHPPPPATTXSTPPP 151

Qy   172 LP------VVLVAPRSTILSMSR 188
           |:         ::||: |
Db   152 PPSATQPAPVAVSPRDTPLLPTR 174


Search completed: March 16, 2006, 16:00:53
Job time : 166 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 16, 2006, 15:58:16 ; Search time 23 Seconds
                                        (without alignments)
                                        263.828 Million cell updates/sec

Title: US-09-989-890-238
Perfect score: 1134
Sequence: 1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 169630 seqs, 28622889 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : Published Applications AA New:*
    1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
    2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
    3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
    4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
    5: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep:*
    6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
    7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
    8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	128.5	11.3	346	7	US-11-036-256-33	Sequence 33, Appl
2	128	11.3	710	7	US-11-072-512-3402	Sequence 3402, Ap
3	127	11.2	246	7	US-11-072-512-2232	Sequence 2232, Ap
4	127	11.2	246	7	US-11-072-512-2243	Sequence 2243, Ap
5	125.5	11.1	240	7	US-11-096-568A-4927	Sequence 4927, Ap
6	125.5	11.1	246	7	US-11-096-568A-4926	Sequence 4926, Ap
7	123	10.8	228	7	US-11-036-256-31	Sequence 31, Appl
8	123	10.8	355	7	US-11-096-568A-23509	Sequence 23509, A
9	123	10.8	723	7	US-11-072-512-2246	Sequence 2246, Ap
10	121.5	10.7	413	7	US-11-096-568A-20771	Sequence 20771, A
11	120.5	10.7	1560	7	US-11-059-982-1	Sequence 1, Appli
12	120	10.6	242	7	US-11-096-568A-26110	Sequence 26110, A
13	120	10.6	375	7	US-11-096-568A-23618	Sequence 23618, A
14	120	10.6	561	6	US-10-467-033-2	Sequence 2, Appli
15	120	10.6	561	6	US-10-467-033-6	Sequence 6, Appli
16	119.5	10.5	244	7	US-11-096-568A-23594	Sequence 23594, A
17	119.5	10.5	411	7	US-11-036-256-89	Sequence 89, Appl
18	118.5	10.4	319	7	US-11-096-568A-24535	Sequence 24535, A
19	118	10.4	681	7	US-11-096-568A-2501	Sequence 2501, Ap
20	118	10.4	1733	7	US-11-182-016-21	Sequence 21, Appl
21	117.5	10.4	904	6	US-10-967-648A-14	Sequence 14, Appl
22	117	10.3	713	6	US-10-330-773-668	Sequence 668, App
23	116	10.2	183	7	US-11-036-256-21	Sequence 21, Appl
24	116	10.2	273	7	US-11-096-568A-22109	Sequence 22109, A
25	116	10.2	652	7	US-11-072-512-3364	Sequence 3364, Ap
26	116	10.2	797	6	US-10-330-773-663	Sequence 663, App
27	115.5	10.2	274	6	US-11-096-568A-22603	Sequence 22603, A
28	115	10.1	395	6	US-10-330-773-661	Sequence 661, App
29	114.5	10.1	312	7	US-11-096-568A-12515	Sequence 12515, A
30	114	10.1	425	7	US-11-096-568A-25808	Sequence 25808, A
31	114	10.1	426	7	US-11-096-568A-25807	Sequence 25807, A
32	114	10.1	430	7	US-11-096-568A-25806	Sequence 25806, A
33	113.5	10.0	245	7	US-11-096-568A-589	Sequence 589, App
34	113.5	10.0	493	7	US-11-096-568A-11657	Sequence 11657, A
35	113	10.0	235	7	US-11-096-568A-24229	Sequence 24229, A
36	113	10.0	319	7	US-11-096-568A-20128	Sequence 20128, A
37	112.5	9.9	252	7	US-11-096-568A-20213	Sequence 20213, A
38	112.5	9.9	256	7	US-11-096-568A-26311	Sequence 26311, A
39	112.5	9.9	286	7	US-11-096-568A-20212	Sequence 20212, A
40	112.5	9.9	306	7	US-11-096-568A-20211	Sequence 20211, A
41	112.5	9.9	459	7	US-11-096-568A-21887	Sequence 21887, A
42	112	9.9	277	7	US-11-096-568A-1871	Sequence 1871, Ap
43	112	9.9	277	7	US-11-096-568A-27199	Sequence 27199, A
44	112	9.9	1336	6	US-10-912-971-10	Sequence 10, Appl
45	111.5	9.8	325	7	US-11-096-568A-24112	Sequence 24112, A

ALIGNMENTS

RESULT 1
US-11-036-256-33
; Sequence 33, Application US/11036256
; Publication No. US20060026719A1
; GENERAL INFORMATION:
; APPLICANT: KIELISZEWSKI, MARCIA
; APPLICANT: XU, JIANFENG
; TITLE OF INVENTION: METHODS OF PRODUCING PEPTIDES/PROTEINS IN PLANTS AND
; TITLE OF INVENTION: PEPTIDES/PROTEINS PRODUCED THEREBY
; FILE REFERENCE: 27211/04130
; CURRENT APPLICATION NUMBER: US/11/036,256
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: 60/602,562
; PRIOR FILING DATE: 2004-08-18
; PRIOR APPLICATION NUMBER: 60/582,027
; PRIOR FILING DATE: 2004-06-22
; PRIOR APPLICATION NUMBER: 60/536,486
; PRIOR FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: Patentin Ver. 3.3
; SEQ ID NO 33
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid construct
US-11-036-256-33

```
Query Match             11.3%;  Score 128.5;  DB 7;  Length 346;
Best Local Similarity   26.6%;  Pred. No. 0.0047;
Matches   53;  Conservative  19;  Mismatches  104;  Indels  23;  Gaps  6;

Qy    2  PHQAAAPVDQTPRTLATMGQRALPSSLALLSRPL------SPPPAACSGDPGCGSGGAGL  54

Db  140  PHSPPPLSPSPTPTPPLGPHSPPPTLSPSPTPTPPPGPHSPPPLSPSPTPTPPLGPHS  199

Qy   55  PSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPAL----IVHPPA  110

Db  200  PPPTLS---PSPTPTPPGPHSPP------PPLSPSPTPTPPLGPHSPPPTLSPSPTPTPPP  252

Qy  111  GGMASGSSQPWAA--ASATPMLSSKASLCIPTRGPPQPLMRTPAARSHWPIPHPCDTAC  168

Db  253  GVPVSGTPLPTLTPLAPTPLLPHPGTPLTLTPAP---TPPLLPHPGTPLPLTLTPL  309

Qy  169  PAPLPVVLVAPRSTILSMS  187

Db  310  PAPTPPLLPHPGTPLPTLT  328
```

RESULT 2
US-11-072-512-3402
; Sequence 3402, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3402
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3402

Query Match   11.3%; Score 128; DB 7; Length 710;
Best Local Similarity   27.5%; Pred. No. 0.01;
Matches   64; Conservative 22; Mismatches 73; Indels 74; Gaps 12;

```
Qy    4 QAAAPVDQTPRTLATMGQRALPSSLALLSRP----------------LSPPPAACSG 44
          ::|||||   |:|||:|||||||||||||||
Db   96 RAPAPVPQ-PFSLPEPSQPILPSVLSLLGLPTPGPSHSDGSFNLLGSDAHLPPPTLSSG 154

Qy   45 D-----------PGCGSGAGLPSASAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPT 92
                     :: |:| || :::|||:|:|:|
Db  155 SPPQPRHPIQPSLPGTTSGS-LSSVPGAPAPPAASKAPV----VPSPVLQSPSEGLGMGA 209

Qy   93 GPRSTMECPPALIVHPPAGGMA--------------SGSSQPWAAASATPM---------- 129
           ::|| :|:   |:|| |
Db  210 GP----ACP----LPPLAGGEAFPFSPSPEQGLALSGAGFP-GMLGALPLPLSLGQPPPSP 260

Qy  130 LSSKASLCIPTRG---PPPQPLMRTPAARSHWPIPHPCDTACPAPLPVVLVAP 179
           |:||||  |:|   :|||::|: |  |:|| |
Db  261 LLNHSLFGVLTGGGGQPPPEPLLPPPGGPG--PPLAPGEPEGPSLLVASLLPP 311
```

RESULT 3
US-11-072-512-2232
; Sequence 2232, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2232
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2232

Query Match   11.2%; Score 127; DB 7; Length 246;
Best Local Similarity   24.9%; Pred. No. 0.0045;
Matches   68; Conservative 17; Mismatches 60; Indels 128; Gaps 14;

```
Qy    7 APVDQTPRTLAT------MGQR-ALPSSLA-LLSRPLSPPPAACSGDPGCGSGAGLPSASA 59
          :|:|  || | |    ||  |  |:||  :| | |:
Db    5 ASVDCTPPSLQTELLPPWGPRSSLPADCAPCLGTALLPPWGPCSSLPPCGPRSSLP----- 60

Qy   60 AAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTM--ECPPAL------------I 105
          ||||| ::  :::: :::|||::|
Db   61 ------GDRAPPS-LRTALG---PPWGPRPSLPADRAPPSLGTSLPPPSLRTV 103

Qy  106 VHPPAGGMASGSSQPWAA--------------------ASATPMLSSKASLC-------- 137
          : ||: ||:
Db  104 LLPPCG---LCFSLPWSSLDLPELCQVGPGRPFPGPAVALSPLTCFHSCCCPRWALGFVLL 160

Qy  138 -----------------IPTR-------------------GPPPQPLMRTPAARSHWPIPH 162
            ||::
Db  161 LKPFVLNLRGSRDWGIPSKELWLGAQKEGHRVLVGCPGSGPPPLPPPSLPASISVQGSPG 220

Qy  163 PCDTACPAPLPVVLVAPRSTILSMSRTWTCRRW 195
          || |
Db  221 PC------------------SPLLSVS----AERW 233
```

RESULT 4
US-11-072-512-2243
; Sequence 2243, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU

```
;   APPLICANT: OTSUKA, MOTOYUKI
;   APPLICANT: NAGAHARI, KENJI
;   APPLICANT: MASUHO, YASUHIKO
;   TITLE OF INVENTION: Novel full length cDNA
;   FILE REFERENCE: 084335-0191
;   CURRENT APPLICATION NUMBER: US/11/072,512
;   CURRENT FILING DATE: 2005-03-07
;   PRIOR APPLICATION NUMBER: US 60/350,978
;   PRIOR FILING DATE: 2002-01-25
;   PRIOR APPLICATION NUMBER: JP 2001-379298
;   PRIOR FILING DATE: 2001-11-05
;   NUMBER OF SEQ ID NOS: 4096
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2243
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-072-512-2243

Query Match             11.2%;    Score 127; DB 7;  Length 246;
Best Local Similarity   24.9%;    Pred. No. 0.0045;
Matches    68; Conservative   17; Mismatches   60;  Indels  128; ; Gaps    14;

Qy       7 APVDQTPRTLAT------MGQR-ALPSSLA-LLSRPLSPPPAACSGDPGCGSGAGLPSASA 59
            ||:|| |.|:|       ||:|: |.||:||  :|:||| || || | | |||
Db       5 ASVDCTPPSLQTBILLPPWGPRSSLPADCAPCLGTALLPPWGPCSSLPPCGPRSSLP---- 60

Qy      60 AAGIASSAVBPVCGDAAPACLLRTPLRGLLKPTGPRSTM---ECPPAL----------I 105
            |            | ::  ||            ::|
Db      61 -----------GDRAPPS-LRTALG---PPWGRPSLPADRAPPSLGTSLLPPSLRTv 103

Qy     106 VHPPAGGMASGSSQPWAA-------------ASATPMLSSKASLC------------- 137
            :|||                           |:|||:||||:||:
Db     104 LLPPCG---LCFSLPWSSLDLPELCQVGPGRPFPGPAVALSFLTCFHSCCPRWALGFVLL 160

Qy     138 -----------IPTR-----------------GPPPQPLMRTPAARSHWPIPH 162
                       |||                  ||:| |:
Db     161 LKPFVLNLRGSRDWGIPSKELWLGAQKEGHRVLVGCPGSGPPPLPPPSLPASISVQGSPG 220

Qy     163 PCDTACPAPLPVLVAPRSTILSMSRTWTCRRW 195
            ::|:
Db     221 PC-----------------SPLLSVS-----ABRW 233

RESULT 5
US-11-096-568A-4927
; Sequence 4927, Application US/11096568A
; Publication No. US20060048240A1
;   GENERAL INFORMATION:
;   APPLICANT: Alexandrov, Nickolai et al.
;   TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
;   TITLE OF INVENTION: Therby
;   FILE REFERENCE: 2750-1592PUS2
;   CURRENT APPLICATION NUMBER: US/11/096,568A
;   CURRENT FILING DATE: 2005-04-01
;   NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4927
;   LENGTH: 240
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(240)
;   OTHER INFORMATION: Ceres Seq. ID no. 14305845
US-11-096-568A-4927

Query Match             11.1%;    Score 125.5; DB 7;  Length 240;
Best Local Similarity   33.2%;    Pred. No. 0.0057;
Matches    65; Conservative   16; Mismatches   86;  Indels  29; ; Gaps    12;

Qy       5 AAAPVDQTPRTLIATMGQRALPSSLALLSRPLSPPPAA--CSGDPGCGSGAGLPSASAAAG 62
            ||:|     |||.::||:|:||||:|||||||:||||   |:|| | || |||:|:||
Db      25 AASPTTSPP--AATTTPFASPATAP--SKPKSPAPVASPTSSSPPASS----PNAATATP 76
```

```
Qy      63 IASS--AVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPAL--IVHPPAGGMASGS- 117
            |||    |                   :||||:   :  |||  || | :|||||
Db      77 PASSPTVASPPSKAAPAPVATPPAATPPAATPPAAT---PPAVTPVSSPPAPVPVSSPP 133

Qy     118 ----SQPWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHW-PIPHCDTACPAP 171
                    |:           ||||: ||  | | |  |    ||   :|||
Db     134 APVPVSSPPALAPTTPAPVVAPSAEVP--APAPKSKCKTKKSKKHTAPAPSPSLLGPPAP 191

Qy     172 LPVVLVAPRSTILSMS 187
            | ||||||:|||:||
Db     192 -PV--GAPGSSQDSMS 204

RESULT 6
US-11-096-568A-4926
; Sequence 4926, Application US/11096568A
; Publication No. US20060048240A1
;   GENERAL INFORMATION:
;   APPLICANT: Alexandrov, Nickolai et al.
;   TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
;   TITLE OF INVENTION: Therby
;   FILE REFERENCE: 2750-1592PUS2
;   CURRENT APPLICATION NUMBER: US/11/096,568A
;   CURRENT FILING DATE: 2005-04-01
;   NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4926
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(246)
;   OTHER INFORMATION: Ceres Seq. ID no. 14305844
US-11-096-568A-4926

Query Match             11.1%;    Score 125.5;   DB 7;  Length 246;
Best Local Similarity   33.2%;    Pred. No. 0.0058;
Matches    65; Conservative   16; Mismatches   86;  Indels  29; ; Gaps    12;

Qy       5 AAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAA--CSGDPGCGSGAGLPSASAAAG 62
            ||:|     |||.::||:|:||||:|||||||:||||   |:|| | || |||:|:||
Db      31 AASPTTSPP--AATTTPFASPATAP--SKPKSPAVASPTSSSPPASS----PNAATATP 82

Qy      63 IASS--AVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPAL--IVHPPAGGMASGS- 117
            |||    |                   :||||:   :  |||  || | :|||||
Db      83 PASSPTVASPPSKAAPAPVATPPAATPPAATPPAAT---PPAVTPVSSPPAPVPVSSPP 139

Qy     118 ----SQPWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHW-PIPHCDTACPAP 171
                    |:           ||||: ||  | | |  |    ||   :|||
Db     140 APVPVSSPPALAPTTPAPVVAPSAEVP--APAPKSKCKTKKSKKHTAPAPSPSLLGPPAP 197

Qy     172 LPVVLVAPRSTILSMS 187
            | ||||||:|||:||
Db     198 -PV--GAPGSSQDSMS 210

RESULT 7
US-11-036-256-31
; Sequence 31, Application US/11036256
; Publication No. US2006002719A1
;   GENERAL INFORMATION:
;   APPLICANT: KIELISZEWSKI, MARCIA
;   APPLICANT: XU, JIANFENG
;   TITLE OF INVENTION: METHODS OF PRODUCING PEPTIDES/PROTEINS IN PLANTS AND
;   TITLE OF INVENTION: PEPTIDES/PROTEINS PRODUCED THERBY
;   FILE REFERENCE: 27211/04130
;   CURRENT APPLICATION NUMBER: US/11/036,256
;   CURRENT FILING DATE: 2005-01-14
;   PRIOR APPLICATION NUMBER: 60/602,562
;   PRIOR FILING DATE: 2004-08-18
;   PRIOR APPLICATION NUMBER: 60/582,027
;   PRIOR FILING DATE: 2004-06-22
```

; PRIOR APPLICATION NUMBER: 60/536,486
; PRIOR FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 3.3
;
; SEQ ID NO 31
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid construct
US-11-036-256-31

Query Match   10.8%;   Score 123; DB 7; Length 228;
Best Local Similarity   27.8%;   Pred. No. 0.0084;
Matches   52; Conservative   17; Mismatches   92; Indels   26; Gaps   8;

Qy   2 PHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAAAA   61
              ::| ||| ::||| :|||
Db   64 PHSPPPLSPSPTPTPPLGPHSPPPTLSPSPTP-TPPPGPHSPPPLS-----PSPTPT-   116

Qy   62 GIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHP-PAGGMASGSSQP 120
              |:||| || ||:|| ||||:   :|||
Db   117 ------PPLGPHSPPPL-SPSPTPTPPPGPHSP---PPPLSPSPTPTPPPLGPHSPPP 164

Qy   121 WAAASATPMLSSKASLC---IPTRGPPPQPLMRTPAARSHWPIPHPCDTACPLPVVLV 177
              |||::| ||:| ::|
Db   165 TLSPSPTPTPPPGPVPVSGTPLPTLTPLAP---TPPLLPHPGTPLPTLTPLAPTPPLLP 221

Qy   178 APRSTIL 184
              |:::
Db   222 HPGRTMV 228

RESULT 8
US-11-096-568A-23509
; Sequence 23509, Application US/11096568A
; Publication No. US2006004824OA1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 23509
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(355)
; OTHER INFORMATION: Ceres Seq. ID no. 12413224
US-11-096-568A-23509

Query Match   10.8%;   Score 123; DB 7; Length 355;
Best Local Similarity   27.6%;   Pred. No. 0.013;
Matches   62; Conservative   20; Mismatches 101; Indels   42; Gaps   9;

Qy   5 AAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPPAACSGDPGCGSGAGLPSASAAAGIA   64
              |||||:|:| ||:::||:|:|:| :::||
Db   46 ATATSSSPTPWCGSTPAASARPASTSTTAR------ARRGTRSCSSRSRPPPASTTPSST   98

Qy   65 SSAVEPVCGDAAPACLLRTPLRGLLKPTGPR----STMECPPALIVHPPAGGMASGSSQP 120
              ::|: | |||||| |||||||||||||||||:
Db   99 STSSTP----------TPPRASSRSSAPRASRSATSSTTPGSGARPRA---ASASSAP   143

Qy   121 WAA-----ASATPMLSSKASLC-IPTRGPPPQPLMR----TPAARSHWPIPHPCDTACPA 170
              ||| ||:|||||:||:| ||||||| |:| || |:|||:|||:|
Db   144 RAAPTAASTSASPSRSRPGTTTPTTPTRPRPTPTRAAAPVTPTAR---PLLLPAATTAPE 200

Qy   171 PLPVVLVAPRSTILSMSRTWT--CRRWAVAPCRABKLMCSSSRS 212

RESULT 9
US-11-072-512-2246
; Sequence 2246, Application US/11072512
; Publication No. US2006002945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2246
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2246

Query Match   10.8%;   Score 123; DB 7; Length 723;
Best Local Similarity   29.8%;   Pred. No. 0.024;
Matches   48; Conservative   12; Mismatches   81; Indels   20; Gaps   5;

Qy   8 PVDQTPRTLATMGQRALPSSL---------ALLSRPLSPPPAACSGDPGCGSGAGLPSA   57
              ||||| | | ::|::| |:|:| || :: | | |:|
Db   91 PYSQMPPPLPTMPPPVLPPSLPPPVMPPALPATVPPPGMPPPVMPPSLPTSVPPPGMPPS 150

Qy   58 SAAAGIASSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGS 117
              :::::
Db   151 LSSAGPPPVLPPPSLFSAGPPPVLPPP---SLSSTAPPPVMPLPPLSSATPPPGIPPPGV 207

Qy   118 SQ---PWAAASATPMLSSKASLCIPTRGPPPQP-LMRTPAA 154
              ::| |:|:| ||::||:|
Db   208 PQGIPPQLTAAPVPPASSSQSQVPEK---PRPALLPTPVS 245

RESULT 10
US-11-096-568A-20771
; Sequence 20771, Application US/11096568A
; Publication No. US2006004824OA1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20771

```
;  LENGTH: 413
;  TYPE: PRT
;  ORGANISM: Zea mays subsp. mays
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(413)
;  OTHER INFORMATION: Ceres Seq. ID no. 12387083
US-11-096-568A-20771

Query Match             10.7%;  Score 121.5; DB 7;  Length 413;
Best Local Similarity   25.3%;  Pred. No. 0.019;
Matches    56;  Conservative  24;  Mismatches  80;  Indels  61;  Gaps  11;

Qy        1 SPHQAAAPVDQTPRTLATMGQRALPSSLALLSRPLSPPP--------AACSGDPGCG---S 50
            ||:||                    :|||||||||||||                     :
Db      129 SPSRTRSSPTSTPTT--------------RPWTHPPRAPTRSAACCPASPPTGSPCS 171

Qy       51 GAGLPSASAAAGIASSAVBPVCGDAAPACLLRTPLRGLLKPTGPRS-----TMECPPALI 105
            : :||:|||  |||||          |||  | |:   ||      :::|
Db      172 SSSPPTSSRRSSCSRSSPSPPPCSPSPPS----TP-------PSATRSPSRACSPSCPASGG 221

Qy      106 VHPPAGGMASGSSQPW--------AAASATPMLSSKAS-LCIPTRGPPPQ-PLMRTPAARS 156
            :|||||||                 ::|||||||  |            :    :
Db      222 GSPDRSSPPSRSSSPTTPSPCSSSASSSPRTTAPASRACSPSSSPSPTCSASSTSASSG 281

Qy      157 HWPIPHPCDTACPAPLPVVLV--------APRSTILSMSRT 189
            ||                       ::| ||::|  |||
Db      282 TSPASSPCSRTTRASPPCARARTSSGGSSSPR--VLSSSRS 320
```

```
RESULT 11
US-11-059-982-1
;  Sequence 1, Application US/11059982
;  Publication No. US2005025507A1
;  GENERAL INFORMATION:
;  APPLICANT: Jenkins, Robert B.
;  APPLICANT: Yang, Ping
;  APPLICANT: Thibodeau, Steve
;  APPLICANT: Wang, Liang
;  APPLICANT: Schaid, Daniel
;  TITLE OF INVENTION: CYTOGENETICALLY DETERMINED DIAGNOSIS AND
;  TITLE OF INVENTION: PROGNOSIS OF PROLIFERATIVE DISORDERS
;  FILE REFERENCE: 07039-505001
;  CURRENT APPLICATION NUMBER: US/11/059,982
;  CURRENT FILING DATE: 2005-02-17
;  PRIOR APPLICATION NUMBER: US 60/545,573
;  PRIOR FILING DATE: 2004-02-17
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
;  LENGTH: 1560
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-11-059-982-1

Query Match             10.6%;  Score 120.0; DB 7;  Length 1560;
Best Local Similarity   26.3%;  Pred. No. 0.077;
Matches    65;  Conservative  17;  Mismatches  84;  Indels  81;  Gaps  13;

Qy        1 SPHQ-------AAAPVDQTPRTLATMGQRAL---PSSLALLSRPLSPPPAACSGDPGCG 49
            :|||                   :||| ||                   ||||||||||
Db      531 APHSGAHSAHILSAAPIQ--------VGQPALFQMPVSLAAGSLPTQSQPAPAG--PAAT 580

Qy       50 S---GAGL-PSASAAAGIASSAVBP-----VCGDAAPACLLR--------TPL-RGLL 89
            :   |||| |                 :: ||||||||||        |||  ||||
Db      581 TVLQGVTLPPSAVAMLNTPDGLVQPATPAAATGRAAPVLTVQPAPQAPPAVSTPLPLGLQ 640

Qy       90 K------PTGPRSTMECPPALIVHPPAGGMASG-------------------- 116
            |       :|:|||
Db      641 QPQAQQPPQAPTPQAAAPPQATTPQPSPGLASSPPEKIVLGQPPSATPTAILTQDSLQMFL 700

Qy      117 ----SSQPWAAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPH-PCDTACPAP 171
            |||  ||      |||||  ||  |||
Db      700
```

```
Db      701 PQERSCQPLSAEG--PHLSVPASVIVSAPPPAQDPAPATPVAKGAGLGPQAPDSQASPAP 758

Qy      172 LPVVLVA 178
            | :||
Db      759 APQIPAA 765
```

```
RESULT 12
US-11-096-568A-26110
;  Sequence 26110, Application US/11096568A
;  Publication No. US20060048240A1
;  GENERAL INFORMATION:
;  APPLICANT: Alexandrov, Nickolai et al.
;  TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
;  TITLE OF INVENTION: Therby
;  FILE REFERENCE: 2750-1592US2
;  CURRENT APPLICATION NUMBER: US/11/096,568A
;  CURRENT FILING DATE: 2005-04-01
;  NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 26110
;  LENGTH: 242
;  TYPE: PRT
;  ORGANISM: Zea mays subsp. mays
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(242)
;  OTHER INFORMATION: Ceres Seq. ID no. 13498595
US-11-096-568A-26110

Query Match             10.6%;  Score 120; DB 7;  Length 242;
Best Local Similarity   29.6%;  Pred. No. 0.015;
Matches    55;  Conservative  19;  Mismatches  70;  Indels  42;  Gaps  8;

Qy       31 LSRPLSPPPAACSGDPGCGSGGAGLPSASAA--AGIASSAVBPVCGDAAPACLLRTPLRGL 88
            ||:||:||                  |:|||  :|  ||||||||||||||||:|||
Db       19 VSPPISVPPGAAGGDGGAEAAAGRRAAGASPQAGRAGAGL-GAPGTSGGCRSRRP----- 73

Qy       89 LKPTGPRSTMECPPALIVHPPAGGMASGSS--QPWAAAASATPMLSSKASLCIPTRGPPPQ 146
                          |||  ||||||||||          ||:|||||||||||||||
Db       74 ---SSASASGCSSASPAPPSCSCSSSSSSASPAAAAA------ASPPTPTTPPTPS 122

Qy      147 PLMRTPAARSHWPIPHPCDTACPAPLPVVLVAPRSTILSMSRTWTCRRWAVAPCRABKLM 206
            ||:|               |||                 :::  |||
Db      123 CSRSTCSSRPR------------PRPRP---PRTSRRSSAASRSPRCSSPRPCR----- 161

Qy      207 CSSSRS 212
            ||||
Db      162 -SRSRS 166
```

```
RESULT 13
US-11-096-568A-23618
;  Sequence 23618, Application US/11096568A
;  Publication No. US20060048240A1
;  GENERAL INFORMATION:
;  APPLICANT: Alexandrov, Nickolai et al.
;  TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
;  TITLE OF INVENTION: Therby
;  FILE REFERENCE: 2750-1592US2
;  CURRENT APPLICATION NUMBER: US/11/096,568A
;  CURRENT FILING DATE: 2005-04-01
;  NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 23618
;  LENGTH: 375
;  TYPE: PRT
;  ORGANISM: Zea mays subsp. mays
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(375)
;  OTHER INFORMATION: Ceres Seq. ID no. 12414168
US-11-096-568A-23618

Query Match             10.6%;  Score 120; DB 7;  Length 375;
```

```
; TITLE OF INVENTION: Mutant SH3-Binding Protein Compositions and Methods
; FILE REFERENCE: H0498.70204US00
; CURRENT APPLICATION NUMBER: US/10/467,033
; CURRENT FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: US 60/266,129
; PRIOR FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 6
; LENGTH: 561
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (415)..(415)
; OTHER INFORMATION: Xaa = any amino acid
; OTHER INFORMATION: provided that the mutant domain is not wild type
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (418)..(418)
; OTHER INFORMATION: Xaa = any amino acid
; OTHER INFORMATION: provided that the mutant domain is not wild type
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (420)..(420)
; OTHER INFORMATION: Xaa = any amino acid
; OTHER INFORMATION: provided that the mutant domain is not wild type
US-10-467-033-6

Query Match             10.6%;   Score 120;  DB 6;  Length 561;
Best Local Similarity   28.8%;   Pred. No. 0.033;
Matches   63;  Conservative  16;  Mismatches 100;  Indels  40;  Gaps   9;

Qy      8  PVDQTPRTLATMGQRALPSSLALLSR--PLSPPPAACSGDPGCGSGAGLPSASAAAGIA 64
           | |||      ||| :  : :| |||        |||:||||:|||:|      |||
Db    207  PPVPTPRKPAFSD--MPRAHSFTSKGPGPLLPPPPKHGLPDVGLAA---------E 252

Qy     65  SSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW--- 121
           : :| |||       ||:||||||||| ||||||||    |:| ||| ||  ::||
Db    253  DSKRDPLCPRRAEPC-PRVPATPRRMSDPPLSTMPTAPGLRKPCFRESASPSPEPWTPG 311

Qy    122  ---AAASATPMLSSKASLC------------IPTRGPPPQPLMRTPAARSHWPIPHPCD 165
              |||||       | |             |||||||  ||    |||   |   |:
Db    312  HGACSTSSAAIMATATSRNCDKLKSFHLSPRGPPTSEPPVPANKPKFLKIAEEDP-PRE 370

Qy    166  TACPAPLPVVLVAPRSTILSMSRTWTCRRWAVAPCRAEK 204
           ||     ||     ||| ||:|||||||||||||
Db    371  AAMPG-LFVPPVAPRPPALKLPVPEAMARPAVLP-RPEK 407
```

```
Best Local Similarity   26.6%;   Pred. No. 0.022;
Matches   63;  Conservative  22;  Mismatches  98;  Indels  54;  Gaps  12;

Qy      5  AAAPVDQTPRTLATMGQRA------LPSSLALLSRPLSPPPAACSGD--PGCGSGAGL 54
           |||               |||      : :| |||     ||||| | |   ||||||
Db    156  AGTPGSPRPPTSAASSCRATRPAAPSRTTSPCASVARPRTWPPSPC-GDTSSSCHSSA-- 212

Qy     55  PSASAAAGIASSAVEPVCGDAAP------ACLLRTPLRGLLKPTGPRSTMECPPALIVHP 108
           | | : | ||                     |||||||||||||   ||||||    |:|
Db    213  --AWRARGRRPSA-PPTLSSTGPSTTATGACRCRRAPR-----FTTPWPTRSAPARRRWTP 265

Qy    109  PAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGPPPQP--------LMRTPAARSHW 158
           ||||||   || ||||||||       | |||||||||||          | ||  :|
Db    266  -----SSRPPWSSSAAATSSTTAPSTTPGPSGPPGSPWSGTSTASSPSTPGPT 319

Qy    159  PIPHPCDTACPAPLPVVLVAPRSTILSMSRTWTCRRWAVAPCRAEKLMC----SSSR 211
           | | : | ||          |||  |:||||||||||||||| ||| :| :
Db    320  PPPSCASSASSTPT--AASTSTDTSVSNS------SIKSEGASRMVCVEPSSSSSR 368
```

```
RESULT 14
US-10-467-033-2
; Sequence 2, Application US/10467033
; Publication No. US2006019248A1
; GENERAL INFORMATION:
; APPLICANT: Tiziani, Valdenize
; APPLICANT: Reichenberger, Ernest
; APPLICANT: Ueki, Yasuyoshi
; APPLICANT: Olsen, Bjorn R.
; TITLE OF INVENTION: Mutant SH3-Binding Protein Compositions and Methods
; FILE REFERENCE: H0498.70204US00
; CURRENT APPLICATION NUMBER: US/10/467,033
; CURRENT FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: US 60/266,129
; PRIOR FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 2
; LENGTH: 561
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-467-033-2

Query Match             10.6%;   Score 120;  DB 6;  Length 561;
Best Local Similarity   28.8%;   Pred. No. 0.033;
Matches   63;  Conservative  16;  Mismatches 100;  Indels  40;  Gaps   9;

Qy      8  PVDQTPRTLATMGQRALPSSLALLSR--PLSPPPAACSGDPGCGSGAGLPSASAAAGIA 64
           | |||      ||| :  : :| |||        |||:||||:|||:|      |||
Db    207  PPVPTPRKPAFSD--MPRAHSFTSKGPGPLLPPPPKHGLPDVGLAA---------E 252

Qy     65  SSAVEPVCGDAAPACLLRTPLRGLLKPTGPRSTMECPPALIVHPPAGGMASGSSQPW--- 121
           : :| |||       ||:||||||||| ||||||||    |:| ||| ||  ::||
Db    253  DSKRDPLCPRRAEPC-PRVPATPRRMSDPPLSTMPTAPGLRKPCFRESASPSPEPWTPG 311

Qy    122  ---AAASATPMLSSKASLC------------IPTRGPPPQPLMRTPAARSHWPIPHPCD 165
              |||||       | |             |||||||  ||    |||   |   |:
Db    312  HGACSTSSAAIMATATSRNCDKLKSFHLSPRGPPTSEPPVPANKPKFLKIAEEDP-PRE 370

Qy    166  TACPAPLPVVLVAPRSTILSMSRTWTCRRWAVAPCRAEK 204
           ||     ||     ||| ||:|||||||||||||
Db    371  AAMPG-LFVPPVAPRPPALKLPVPEAMARPAVLP-RPEK 407
```

```
RESULT 15
US-10-467-033-6
; Sequence 6, Application US/10467033
; Publication No. US2006019248A1
; GENERAL INFORMATION:
; APPLICANT: Tiziani, Valdenize
; APPLICANT: Reichenberger, Ernest
; APPLICANT: Ueki, Yasuyoshi
; APPLICANT: Olsen, Bjorn R.
```

Search completed: March 16, 2006, 16:01:22
Job time : 24 secs

```
                    GenCore version 5.1.7
         Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      March 17, 2006, 03:55:30 ; Search time 7434 Seconds
                              (without alignments)
                              1621.040 Million cell updates/sec

Title:         US-09-989-890-238
Perfect score: 1134
Sequence:      1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table: BLOSUM62
               Xgapop  10.0 , Xgapext  0.5
               Ygapop  10.0 , Ygapext  0.5
               Fgapop   6.0 , Fgapext  7.0
               Delop    6.0 , Delext   7.0

Searched:      5883141 seqs, 2842172553 residues

Total number of hits satisfying chosen parameters:    11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095246_16568/app_query.fasta_1
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss04
-USER=US09989890 @CGN_1 1 4939 @runat 16032006_095246 16568 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database :   GenEmbl:*

```
              1: gb_ba:*
              2: gb_in:*
              3: gb_env:*
              4: gb_om:*
              5: gb_ov:*
              6: gb_pat:*
              7: gb_ph:*
              8: gb_pr:*
              9: gb_ro:*
             10: gb_sts:*
             11: gb_sy:*
             12: gb_un:*
             13: gb_vi:*
             14: gb_htg:*
             15: gb_pl:*
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1134	100.0	1197	6	CQ717675	CQ717675 Sequence
2	1134	100.0	1679	6	BC033143	BC033143 Homo sapi
3	1134	100.0	1713	6	AX644347	AX644347 Sequence

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

4	1134	100.0	140207	8	AL356390	AL356390	Human DNA
5	1128	99.5	1785	8	AX747274	AX747274	Sequence
6	1128	99.5	1785	8	AK091952	AK091952	Homo sapi
7	1126	99.5	111738	14	AL592304	AL592304	Homo sapi
9	1002.5	88.4	1977	6	AX535029	AX535029	Sequence
9	1002.5	88.4	1977	6	AX535090	AX535090	Sequence
10	685	60.4	193813	9	AL627228	AL627228	Mouse DNA
11	680	60.0	1680	9	BC006890	BC006890	Mus muscu
12	653	57.6	1718	9	BC079148	BC079148	Rattus no
13	653	57.6	235419	14	AC095979	AC095979	Rattus no
c 14	653	57.6	249406	14	AC119963	AC119963	Rattus no
c 15	653	57.6	255017	14	AC126312	AC126312	Rattus no
c 16	417	36.8	427	1	AX071660	AX071660	Sequence
c 17	186	16.4	348764	1	BX569689	BX569689	Synechoco
18	184.5	16.3	192533	8	AC055876	AC055876	Homo sapi
c 19	182	16.0	216408	14	AC092466	AC092466	Homo sapi
c 20	182	16.0	222605	8	AC010973	AC010973	Homo sapi
c 21	178.5	15.7	110000	1	BA000030_17	BA000030_17	Continuation (18 o
c 22	178	15.7	1716	6	AR627633	AR627633	Sequence
c 23	178	15.7	9198	6	AR619571	AR619571	Sequence
c 24	176	15.5	58734	15	AC146481	AC146481	Oryza sat
c 25	176	15.5	110000	15	AP008216_180	AP008216_180	Continuation (181
c 26	176	15.5	247875	14	AC137235	AC137235	Rattus no
c 27	174	15.3	86571	5	AC151463	AC151463	Xenopus t
28	173.5	15.3	42149	1	AJ786317	AJ786317	Streptomy
29	173	15.3	148845	8	AP001067	AP001067	Homo sapi
30	173	15.3	187114	14	AC026998	AC026998	Homo sapi
31	173	15.3	340000	8	AP001754	AP001754	Homo sapi
c 32	172	15.2	250480	14	AC115644	AC115644	Rattus no
c 33	170	15.0	41932	8	AL158091	AL158091	Human DNA
c 34	169.5	14.9	110000	8	BA000030_82	BA000030_82	Continuation (83 o
c 35	169	14.9	864	2	AY819649	AY819649	Eimeria t
c 36	169	14.9	864	6	I08369	I08369	Sequence 2
c 37	169	14.9	864	6	I09629	I09629	Sequence 2
c 38	169	14.9	6567	6	I09636	I09636	Sequence 16
c 39	169	14.9	7053	2	ETB306453	ETB306453	Eimeria t
c 40	168.5	14.9	672	6	I09630	I09630	Sequence'4
c 41	168.5	14.9	2703	6	AR621538	AR621538	Sequence
c 42	168.5	14.9	23417	6	AR619812	AR619812	Sequence
c 43	168	14.8	1761	6	CQ850808	CQ850808	Sequence
c 44	168	14.8	1761	6	AK128822	AK128822	Homo sapi
c 45	168	14.8	299050	1	SCO939119	SCO939119	Streptomy

ALIGNMENTS

```
RESULT 1
CQ717675
LOCUS          CQ717675                 1197 bp    DNA            linear   PAT 03-FEB-2004
DEFINITION     Sequence 3609 from Patent WO02068579.
ACCESSION      CQ717675
VERSION        CQ717675.1  GI:4227B532
KEYWORDS       .
SOURCE         Homo sapiens (human)
  ORGANISM     Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1   (bases 1 to 1197)
  AUTHORS      Venter,C.J., Adams,M.C., Li,P.W. and Myers,B.W.
  TITLE        Kits, such as nucleic acid arrays, comprising a majority of
               humanexons or transcripts, for detecting expression and other uses
               thereof
  JOURNAL      Patent: WO 0206B579-A 3609 06-SEP-2002;
               PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1..1197
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
```

```
Pred. No.:
Score:                4.37e-26
Percent Similarity:   1134.00
Best Local Similarity: 100.0%
Query Match:          100.0%
DB:

Length:    1197
Matches:   212
Conservative:  0
Mismatches:    0
Indels:        0
Gaps:          6

US-09-989-890-238 (1-212) x CQ717675 (1-1197)

Qy     1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly   20
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   104  AGCCCCCACCAGCCGCGCACCCGTAGACCAGACCCCAAGGACCCTGGCACCATGGGC  163

Qy    21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla   40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   164  CAGAGAGCAGCATTACCTTCTCTGGCCTCTGCTGCTGAGCCGGCCCCTTGAGTCCCCACCTGCT  223

Qy    41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   224  GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCCGGGCTGCCTTCGCGCCGCT  283

Qy    61  AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu   80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   284  GCCGGGATTGCCTCCAGCGGCTGTGGAGCCTGTGCGGGGATGCAGCCCCTGCCTGTCTA  343

Qy    81  LeuArgThrProLeuArgGlyLeuLeuLysProThrHisGlyProArgSerThrMetGluCys  100
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   344  CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC  403

Qy   101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnInPro  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404  CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCCAGCGGCTCAAGTCAACCA  463

Qy   121  TrpAlaAlaAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   464  TGGGCAGCAGCTTCAGTAACCCCGATGTTAAGCTCAAAGGCATCCCTGTATCCCTACC  523

Qy   141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   524  CGAGGGGCCACCTCCCCAGCCCCTGATGCGGGACTCCTGCTGCAAGGAGCCACTGGCCGATC  583

Qy   161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg  180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   584  CCCACCCCATGCCAGACACACCCTGCCCAGCACCTTTGCCCAGTAGTCCTCCTGGCTCCGAGG  643

Qy   181  SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys  200
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   644  AGTACTATTCTTTTCCATGAGTCGGACCTGGACCTGCCCGGAGATGGGGCAGTGGCTCCATGT  703

Qy   201  ArgAlaGluLysMetCysSerSerSerArgSer   212
             ||||||||||||||||||||||||||||||||||
Db   704  CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAGAGC  739

RESULT 2
BC033143
LOCUS      BC033143                1679 bp    mRNA    linear   PRI 03-AUG-2005
DEFINITION Homo sapiens chromosome 1 open reading frame 172, mRNA (cDNA clone
           MGC:45763 IMAGE:3960220), complete cds.
ACCESSION  BC033143
VERSION    BC033143.1  GI:21619958
KEYWORDS   MGC.
SOURCE     Homo sapiens (human)
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Hominidae; Homo.
REFERENCE  1  (bases 1 to 1679)
  AUTHORS  Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D.,
           Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
           Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
           Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,J.,
           Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
           Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
```

(continuation of reference/authors — right column)

```
           Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
           Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
           Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
           McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
           Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
           Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
           Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
           Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
           Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
           Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
           Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
           Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

CONSRTM    Mammalian Gene Collection Program Team
  TITLE    Generation and initial analysis of more than 15,000 full-length
           human and mouse cDNA sequences
  JOURNAL  Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED  12477932
REFERENCE  2  (bases 1 to 1679)
  CONSRTM  NIH MGC Project
  AUTHORS  Direct Submission
  TITLE    Submitted (25-JUN-2002) National Institutes of Health, Mammalian
  JOURNAL  Gene Collection (MGC), Bethesda, MD 20892-2590, USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Genome Sequence Centre,
           BC Cancer Agency, Vancouver, BC, Canada
           info@bcgsc.bc.ca
           Martin Hirst, Thomas Zeng, Ryan Morin, Michelle Moksa, Johnson
           Pang, Diana Mah, Jing Wang, Kieth Fichter, Eric Chuah, Allen
           Delaney, Rob Kirkpatrick, Agnes Baross, Sarah Barber, Mabel
           Brown-John, Steve S. Chand, William Chow, Ryan Babakaiff, Dave
           Wong, Corey Matsuo, Jaclyn Beland, Susan Gibson, Luis delRio, Ruth
           Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
           Kim MacDonald, Mike R. Mayo, Josh Moran, Diana Palmquist, JR
           Santos, Duane Smailus, Jeff Stott, Miranda Tsai, George Yang,
           Jacquie Schein, Asim Siddiqui, Steven Jones, Rob Holt, Marco Marra.

           Clone distribution: MGC clone distribution information can be found
           through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
           Series: IRAL Plate: 43 Row: c Column: 17.

FEATURES             Location/Qualifiers
     source          1..1679
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:45763 IMAGE:3960220"
                     /tissue_type="Placenta, choriocarcinoma"
                     /clone_lib="NIH_MGC_21"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1..1679
                     /gene="C1orf172"
                     /note="synonyms: FLJ34633, RP11-344H11.3"
                     /db_xref="GeneID:126695"
     CDS             387..1121
                     /gene="C1orf172"
                     /codon_start=1
                     /product="C1orf172 protein"
                     /protein_id="AAH33143.1"
                     /db_xref="GI:21619959"
                     /db_xref="GeneID:126695"
                     /translation="MGSSFSYPDVKLKGIPVYPYPRATSPAPDADSCCKEPLADPPPM
                     RHSLPSTFASSPRGSEBYYSFPHESDLLPEMGSGSGMSSREIDVLIFKKLTELFSVHQI
                     DELAKCTSDTVFLEKTSKISDLISSITQDYHLDEQDAGRLVRGIIRISTRKSRARPQ
                     TSEGRSTRAAAPTAAAPDSGHETMVGSGLSQDELTVQISQETTADAIARKLRPYGAPG
                     YPASHDSSFPQGTDTDSSGAPLLQVYC"

ORIGIN
```

RESULT 3

```
Alignment Scores:
Pred. No.:
Score:                      5.74e-26        Length:              1679
Percent Similarity:         1134.00         Matches:             212
Best Local Similarity:      100.0%          Conservative:        0
Query Match:                100.0%          Mismatches:          0
DB:                         100.0%          Indels:              0
                                            Gaps:                8
```

US-09-989-890-238 (1-212) x BC033143 (1-1679)

```
Qy    1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly   20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28  AGCCCCCACCAGCAGGCCGCCGCACCCGTAGACCAGGACCCCAAGGAACCCCGGGC       87

Qy   21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerLeuSerLeuArgProLeuSerProLeuProProAla   40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   88  CAGAGAGCATTACCTTCATCTCGGCTCGTCCTGGAGCCGGCCCTTGAGTCCCCCACCTGCT  147

Qy   41  AlaCysSerGlyAspProGlyCysGlyAlaGlyLeuLeuGlyProSerAlaSerAlaAla    60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  148  GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGCCCTTCGTCTTCCGCCGCT      207

Qy   61  AlaGlyIleAlaSerSerAlaValAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu   80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  208  GCCGGGATTGCCTCCAGCGCTGTGGCGGTGGAGCCTGTGTGCGGGGATGCAGCC         267

Qy   81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys   100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  268  CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGACACAATGGAGTGC    327

Qy  101  ProProAlaLeuIleValHisProProAlaGlySerGlyMetAlaSerGlyGlnPro   120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  328  CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAGCGGCTCAAGTCAACCA  387

Qy  121  TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLeuLysAlaSerLeuCysIleProThr  140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  388  TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCGTGTATCCCTACC    447

Qy  141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle   160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  448  CGAGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTCGCAAGGAGCCACTGGCCGATC   507

Qy  161  ProHisProCysAspThrAlaCysProLeuProValValLeuValAlaProArg   180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  508  CCCCACCCATGCGGACACAGCCTGCCCAGCACCTTTGCCAGTAGTCCTCGTGGCTCCGAGG  567

Qy  181  SerThrIleLeuSerMetSerSerArgThrThrCysThrArgArgTrpAlaValAlaProCys   200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  568  AGTACTATTCTTTCCATGAGTCGGACCTGGACCTGGCGCCGGAGATGGGCAGTGGCTCCATGT  627

Qy  201  ArgAlaGluLysLeuMetCysSerSerSerArgSer  212
              |||||||||||||||||||||||||||||||||||||
Db  628  CGAGCCCGAGAAATTGATGTGCTCATCTTCAAGAGC  663
```

RESULT 3
```
AX644347
LOCUS       AX644347              1713 bp    DNA     linear   PAT 27-FEB-2003
DEFINITION  Sequence 105 from Patent WO02077232.
ACCESSION   AX644347
VERSION     AX644347.1  GI:28610406
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Salceda,S., Macina,R.A., Recipon,H., Pluta,J., Sun,Y. and Liu,C.
  TITLE     Compositions and methods relating to breast specific genes and
            proteins
  JOURNAL   Patent: WO 02077232-A 105 03-OCT-2002;
            Diadexus, Inc. (US)
FEATURES             Location/Qualifiers
```

(middle column)

```
source          1..1713
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.:
Score:                      5.83e-26        Length:              1713
Percent Similarity:         1134.00         Matches:             212
Best Local Similarity:      100.0%          Conservative:        0
Query Match:                100.0%          Mismatches:          0
DB:                         100.0%          Indels:              0
                                            Gaps:                6
```

US-09-989-890-238 (1-212) x AX644347 (1-1713)

```
Qy    1  SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly   20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104  AGCCCCCACCAAGCCGCCGCACCCGTAGACCAGGACCCCAAGGAACCCCGGCCACCATGGGC  163

Qy   21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerLeuSerLeuArgProLeuSerProProAla   40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164  CAGAGAGCATTACCTTCATCTCGGCTCGTCCTGGAGCCGGCCCTTGAGTCCCCCACCTGCT   223

Qy   41  AlaCysSerGlyAspProGlyCysGlySerGlyLeuLeuGlyProSerAlaSerAlaAla    60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  224  GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGCCCTGCCTTCCGCCGCT        283

Qy   61  AlaGlyIleAlaSerSerAlaValAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu   80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  284  GCCGGGATTGCCTCCAGCGCTGTGGCGGTGGAGCCTGTGTGCGGGGATGCAGCC         343

Qy   81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys   100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  344  CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGACACAATGGAGTGC    403

Qy  101  ProProAlaLeuIleValHisProProAlaGlyThrGlyMetAlaSerGlySerGlnPro   120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  404  CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAGCGGCTCAAGTCAACCA  463

Qy  121  TrpAlaAlaAlaSerAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  464  TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCGTGTATCCCTACC    523

Qy  141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIle   160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  524  CGAGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTCGCAAGGAGCCACTGGCCGATC   583

Qy  161  ProHisProCysAspThrAlaCysProLeuProValValLeuValAlaProArg   180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  584  CCCCACCCATGCGGACACAGCCTGCCCAGCACCTTTGCCAGTAGTCCTCGTGGCTCCGAGG  643

Qy  181  SerThrIleLeuSerMetSerSerArgThrThrCysThrArgArgTrpArgThrAlaValAlaProCys   200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  644  AGTACTATTCTTTCCATGAGTCGGACCTGGACCTGCGCCGGAGATGTGGCTCCATGT     703

Qy  201  ArgAlaGluLysLeuMetCysSerSerSerArgSer  212
              |||||||||||||||||||||||||||||||||||||
Db  704  CGAGCCCGAGAAATTGATGTGCTCATCTTCAAGAAGC  739
```

RESULT 4

```
AL356390
LOCUS       AL356390            140207 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-344H11 on chromosome 1 Contains
            the NROB2 gene for nuclear receptor subfamily 0 (group B, member
            2), the 5' end of a novel gene (FLJ12455), the NUDC gene formercul
            distribution gene C homolog (A. nidulans), a novel gene (FLJ34633),
            a ribosomal protein L12 (RPL12) pseudogene, a ribosomal protein L32
            (RPL32) pseudogene, a novel pseudogene (DC2), two novel genes and
            two CpG islands, complete sequence.
ACCESSION   AL356390
VERSION     AL356390.24  GI:21436506
KEYWORDS    HTG; DC2; FLJ12455; FLJ34633; NROB2; NUDC; RPL12; RPL32.
```

```
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 140207)
AUTHORS     Glithero,R.
TITLE       Direct Submission
JOURNAL     Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK, E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 17, 2002 this sequence version replaced gi:20218554.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-344H11 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            -------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1..140207
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-344H11"
                     /clone_lib="RPCI-11.2"
     misc_feature    496..726
                     /note="Sequence from overlapping clone RP4-633N17
                     (AL137860). Assembly confirmed by restriction digest."
     misc_feature    898..1641
                     /note="Sequence from overlapping clone RP4-633N17
                     (AL137860). Assembly confirmed by restriction digest."
     misc_feature    2000
                     /note="Clone_right_end: RP4-633N17"
     gene            join(21468..22032,27482..28803,29251..29417)
                     /gene="RP11-344H11.8"
                     /locus_tag="RP11-344H11.8-002"
     mRNA            join(21468..22032,27482..28803,29251..29417)
                     /gene="RP11-344H11.8"
                     /locus_tag="RP11-344H11.8-002"
                     /product="novel protein"
                     /note="match: cDNAs: Em:AK122816.1"
     gene            join(21603..22032,27482..29419)
                     /gene="RP11-344H11.8"
                     /locus_tag="RP11-344H11.8-001"
     mRNA            join(21603..22032,27482..29419)
                     /gene="RP11-344H11.8"
                     /locus_tag="RP11-344H11.8-001"
                     /product="novel protein"
                     /note="match: cDNAs: Em:AK028944.1 Em:BC012790.1
                     Em:BC012790.2 Em:BC014160.1 Em:BC014160.2 Em:BC038001.1"
     CDS             join(21772..22032,27482..28495)
                     /gene="RP11-344H11.8"
```

```
                     /locus_tag="RP11-344H11.8-001"
                     /standard_name="OTTHUMP00000004474"
                     /note="match: proteins: Tr:Q8CI52 Tr:Q8CI10 Tr:Q96A09"
                     /codon_start=1
                     /product="novel protein"
                     /protein_id="CAI13555.1"
                     /db_xref="GI:55959148"
                     /db_xref="Genew:28273"
                     /db_xref="UniProt/TrEMBL:Q96A09"
                     /translation="MPSESGAERRDRAAAQVGTAAATAVATARAPAGGGPDPEALSAFP
                     GRHLSGLSWPQVKRLDALLSEPIPIHGRGNFPTLSVQPRQIVQVRSTLEEQGLHVHS
                     VRLHGSAASHVLHPESGLGYKDLDLVFPRVDLRSEASFQLTKAVVLACLLDFLPAGVSR
                     AKITPLTLKEAVVQKLVKVCTDSDRWSLISLSNKSGKNVELKPVDSVRRQFEFSIDSF
                     QIILDSLLLFGQCSSTPMSEAFHPTVTGESLYGDFTEALEHLRHRVIATRSPEIRGG
                     GLLKYCHLLVRGFPRPSTDVRALQRYMCSRFFIDFPDLVEQRRTLERYLEAHFGGAD
                     AARRYACLVTLRHVVNESTVCLMNHERRQTLDLIAALALQALAEQGPAATAALAWRPP
                     GTDGVVPATVNYTVTPVQPLLAHAYPTWLPCN"
     polyA_signal    29398..29403
                     /gene="RP11-344H11.8"
     polyA_site      29419
                     /gene="RP11-344H11.8"
     gene            complement(join(33541..34592,39827..40731))
                     /gene="RP11-344H11.8-001"
                     /locus_tag="RP11-344H11.7-001"
     mRNA            complement(join(33541..34592,39827..40731))
                     /gene="RP11-344H11.7-001"
                     /product="novel transcript"
                     /note="match: ESTs: Em:AI435085.1 Em:BG752663.1
                     Em:BM021923.1 Em:BM682682.1 Em:BQ646324.1 Em:BQ892187.1
                     Em:CA438061.1"
     gene            48099..48542
                     /gene="RP11-344H11.6-001"
                     /pseudo
     CDS             48099..48542
                     /locus_tag="RP11-344H11.6-001"
                     /note="match: proteins: Tr:AAH54857 Tr:BAB28595
                     Tr:EAA12661 Tr:Q7ZWJ3 Tr:Q8TBU1 Tr:Q9CPZ2 Tr:Q9NRP0
                     Tr:Q9P075 Tr:Q9P1R4"
                     /pseudo
                     /codon_start=1
                     /product="novel pseudogene"
                     /db_xref="PSEUDO:CAI13556.1"
     gene            complement(50415..50811)
                     /locus_tag="RP11-344H11.5-001"
                     /pseudo
     CDS             complement(50415..50811)
                     /locus_tag="RP11-344H11.5-001"
                     /note="match: proteins: Sw:O94008 Sw:P02433 Sw:P17932
                     Sw:P38061 Tr:AAH46339 Tr:AAP80706 Tr:BAC21646 Tr:BAC25812
                     Tr:CAD98375 Tr:O60373 Tr:Q29280 Tr:Q86JP7 Tr:Q86QS0
                     Tr:Q8I7D3 Tr:Q8UVG1 Tr:Q90YT6 Tr:Q9TTX8"
                     /pseudo
                     /codon_start=1
                     /product="ribosomal protein L32 (RPL32) pseudogene"
     gene            complement(join(53464..53610,53925..54274))
                     /locus_tag="RP11-344H11.4-001"
                     /pseudo
     CDS             complement(join(53464..53610,53925..54274))
                     /locus_tag="RP11-344H11.4-001"
                     /note="match: proteins: Sw:P23358 Sw:P30050 Sw:P35979
                     Tr:AAM68299 Tr:CAA16156 Tr:EAA13967 Tr:O60886 Tr:Q7ZUG1
                     Tr:Q862X1 Tr:Q8AVW0 Tr:Q8C2K0 Tr:Q90YV6 Tr:Q9NQO2"
                     /pseudo
                     /codon_start=1
                     /product="ribosomal protein L12 (RPL12) pseudogene"
     gene            74033..84877
                     /gene="RP11-344H11.3"
     mRNA            join(74033..74089,82027..83097,83683..83757,84288..84877)
                     /gene="RP11-344H11.3-001"
                     /locus_tag="RP11-344H11.3-001"
                     /product="novel protein"
```

```
misc_feature    /note="match: cDNAs: Em:AK007551.1 Em:AK091952.1
                Em:AX747274.1 Em:BC006890.1 Em:BC033143.1"
                79656..79827
                /gene="RP11-344H11.3"
                /locus_tag="RP11-344H11.3-001"
                /note="Sequence from uni-directional dGTP big dye
                terminator reads only."
CDS             join(82059..83097,83683..83757,84288..84370)
                /gene="RP11-344H11.3"
                /locus_tag="RP11-344H11.3-001"
                /standard_name="OTTHUMP00000004472"
                /note="match: proteins: Tr:O8C1Q5 Tr:Q8N0S7 Tr:Q8NAX2"
                /codon_start=1
                /product="novel protein"
                /protein_id="CAI13559.1"
                /db_xref="GI:55959149"
                /db_xref="UniProt/TrEMBL:Q50P32"
                /translation="MPRPGHPRPASGPPRLGPWERPTELCLESTYDKPPQPPSRRTTR
                PDPKDPGHHGPBSITFISGSAEPALESPTCCLLWRPWVWEWCRAAFCFRRCRDCLQRC
                GACVRGCSPCLSTBDSTEGTARANWAKEHNGVPPSPDRAPPSRRDGQRLKSTMGSSFS
                YPDVKLKGIPVYPYPRATSPAPDADSCCKEPLADPPMRHSLPSTFASSPRGSEEYYS
                FHESDLDLPEMGSGSMSSREIDVLIFFKLTELFSVHQIDELAKCTSDTVFLEKTSKIS
                DLISSITQDYHLDEQDAEGRLVRGIIRISTRKSRARPOTSEGRSTRAAAPTAAAPDSG
                HETMVGSGLSQDELTVQISQETTADAIARKLRPYGAPGYPASHDSSFQGTDTDSSGAP
                LLQVYC"
polyA_site      84877
                /gene="RP11-344H11.3"
                /locus_tag="RP11-344H11.3-001"
gene            complement(join(87577..88309,88753..88871,88966..89049,
                91374..91568,91663..91779,92621..92686,92779..92982,
                110273..110350,112510..112713))
                /gene="NUDC"
mRNA            complement(join(87577..88309,88753..88871,88966..89049,
                91374..91568,91663..91779,92621..92686,92779..92982,
                110273..110350,112510..112713))
                /gene="NUDC"
                /product="nuclear distribution gene C homolog (A.
                nidulans)"
                /note="match: ESTs: Em:BE798032.1
                match: cDNAs: Em:AK012321.1 Em:AL136725.1 Em:BC007280.1
                Em:BC015153.1"
```

Alignment Scores:

Pred. No.:		Length:	140207
Score:	2.0le-24	Matches:	212
Percent Similarity:	1134.00	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	100.0%	Gaps:	8

US-09-989-890-238 (1-212) x AL355390 (1-140207)

```
Qy      1 SerProHisGlnAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly     20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82162 AGCCCCCACCAGCCGCCGCACCCGTCGACCAGACCCCAAGGACCCTGGCCACCATGGGC  82221

Qy     21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla     40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82222 CAGAGAGCATTACCTTCATCTCTGGCTCTGCGAGCCGGCCCCTTGAGTCCCCACCTGCT  82281

Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla     60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82282 GCCTGCTCTGGCGACCCTGGGTGTGGAAGTGGTGCCGGCCTGCCGTCCGCTTCCGCCGCT  82341

Qy     61 AlaGlyIleAlaSerSerAlaValAlaGluProValCysGlyAspAlaAlaProAlaCysLeu    80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82342 GCCGGGATTGCCTCCAGCGTGTGGGGATGCCGTCGCGGGGATGCAGCCCCTGCCTGTCTA  82401

Qy     81 LeuArgThrProLeuArgGlyLeuLeuLysProThrArgGlyProArgSerThrMetGluCys    100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82402 CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC  82461
```

```
Qy    101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPro    120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82462 CCCCAGCCCTGATCGTGCACCCCCAGCCGCGGGATGGCCAGCGGCTCAAGTCAACCA    82521

Qy    121 TrpAlaAlaLeuAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr   140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82522 TGGGCAGCAGCTTCAGCTACCCGATGTTAAGCTCAAAGCCATCCTGTGTATCCCTACC    82581

Qy    141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle     160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82582 CGAGGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCCAAGGACCCACTGGCCGGATC    82641

Qy    161 ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg   180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82642 CCCCACCATGCGACACAGCCTGCCCAGCACCTTTGCCAGTAGTCTTCGTGGCTCCGAGG    82701

Qy    181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaArgArgTrpAlaValAlaAlaProCys   200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82702 AGTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCCGGAGATGGGGCAGTGGCTGCTCCATGT    82761

Qy    201 ArgAlaGlyLysLeuMetCysSerSerSerArgSer     212
            |||||||||||||||||||||||||||||||||||||
Db  82762 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAGC    82797
```

```
RESULT 5
AX747274
LOCUS       AX747274                1785 bp    mRNA       linear      PAT 20-JUN-2003
DEFINITION  Sequence 799 from Patent EP1308459.
ACCESSION   AX747274
VERSION     AX747274.1  GI:32131662
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
  TITLE     Full-length cDNA sequences
  JOURNAL   Patent: EP 1308459-A 799 07-MAY-2003;
            Helix Research Institute (JP) ; Research Association for
            Biotechnology (JP)
FEATURES    Location/Qualifiers
     source  1..1785
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
ORIGIN
```

Alignment Scores:

Pred. No.:		Length:	1785
Score:	9.08e-26	Matches:	211
Percent Similarity:	1128.00	Conservative:	1
Best Local Similarity:	99.5%	Mismatches:	0
Query Match:	99.5%	Indels:	0
DB:	99.5%	Gaps:	0

US-09-989-890-238 (1-212) x AX747274 (1-1785)

```
Qy      1 SerProHisGlnAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly     20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 AGCCCCCACCAGCCGCCGCACCCGTCGACCAGACCCCAAGGACCCTGGCCACCATGGGC    252

Qy     21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla     40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 CAGAGAGCATTACCTTCATCTCTGGCTCTGCGAGCCGGCCCCTTGAGTCCCCACCTGCT    312

Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla     60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  313  GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCGCTTCCGCCGCT  372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61  AlaGlyIleAlaSerSerAlaAlaValGluProValCysGlyGlyAspAlaAlaCysLeu   80

Db  373  GCGGGATTGCCTCCAGCGCCTGTGTGCGGGGATGCAGCCCCTGCCTGTCTA  432
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81  LeuArgThrTrpProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100

Db  433  CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGCCCAAGGAGCACAATGGAGTGC  492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  101  ProProAlaLeuLeuValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro  120

Db  493  CCCCAGCCCTGATCGTGTCGCACCCCCCAGCGCCGGCGGATGGCCGACAGCTCAAGTCAACCA  552
              ||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121  TrpAlaAlaSerAlaThrProMetLeuSerSerLysSerAlaSerLeuCysIleProThr  140

Db  553  TGGGCAGCAGCTTCAGCTACCCCGGATGTTAAGCTCAAAGGCATCCCTGTATCCCTACC  612
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProProIle  160

Db  613  CGAGGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGATC  672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  161  ProHisProCysAspThrAlaCysProLeuProLeuProValValLeuValAlaProArg  180

Db  673  CCCCACCCATGGGACACAGCCGTGCCCCAGCACCTTGCCAGTAGTCCTCGTGGCTCCGAGG  732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  181  SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys  200

Db  733  AGTACTATTCTTTCCATGAGTCGGACCTCGGACCTGCCCGAGATGGGCAGTGGCTCCATGT  792
              |||||||||||||||||||||||||||||||||||||||||||||
Qy  201  ArgAlaGluGluLysMetCysSerSerSerArgSer  212

Db  793  CGAGCCAGAGAAATTGATGTGCTCATCTTCAAGAAGC  828
```

RESULT 6
AK091952

LOCUS   AK091952   1785 bp   mRNA   linear   PRI 30-JAN-2004
DEFINITION Homo sapiens cDNA FLJ34633 fis, clone KIDNE2015710.
ACCESSION   AK091952
VERSION   AK091952.1 GI:21750433
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
AUTHORS   Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
            Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
            Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
            Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
            Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
            Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
            Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
            Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
            Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
            Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
            Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
            Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,
            Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
            Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
            Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
            Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
            Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
            Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
            Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
            Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
            Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
            Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
            Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
            Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
            Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
            Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
            Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
TITLE   Complete sequencing and characterization of 21,243 full-length
            human cDNAs
JOURNAL   Nat. Genet. 36 (1), 40-45 (2004)
PUBMED   14702039
REFERENCE   2
AUTHORS   Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
            Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 1785)
AUTHORS   Isogai,T. and Yamamoto,J.
TITLE   Direct Submission
JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES   Location/Qualifiers
     source   1..1785
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="KIDNE2015710"
                     /tissue_type="kidney"
                     /clone_lib="KIDNE2"
                     /note="cloning vector: pME18SFL3"
     CDS   90..1286
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="BAC03775.1"
                     /db_xref="GI:21750434"
                     /translation="MPRRGHRPRPASGPPRLGPWERPTELCLETYDKPQPPSRRTRR
                     PDPKDPGHHGPESITTFISGSAEPALESPTCCLLWRPWVWEWCRAAFCFRRCRDCLQRC
                     GACVRGCSPCLSTBDSTEGTAEANWAKEHNGVPSPDRAPPSRRDGQQLKSTMGSSFS
                     YPDVKLKGIPVPYPRATSPAPDADSCCKEPLADPPMRHSLPSTFASSPRGSEEYYS
                     FHESDLLPEMGSGSMSSREIDVLIFKKLTELFSVHQIDELAKCTSDTVFLEKTSKIS
                     DLISSITQDYHLDEQDAEGRLVRGIIRISTRKSRARPQTSEGRSTRAAPTAAAPDSG
                     HETMVGSLSQDELTVQISQETTADAIIARKLRPYGAPGYPASHDSSFQGTDTDSSGAP
                     LLQVYC"
ORIGIN

Alignment Scores:

Pred. No.:		Length:	1785
Score:	9.08e-26	Matches:	211
	1128.00		
Percent Similarity:	99.5%	Conservative:	0
Best Local Similarity:	99.5%	Mismatches:	1
Query Match:	99.5%	Indels:	0
DB:	8	Gaps:	0

US-09-989-890-238 (1-212) x AK091952 (1-1785)

```
Qy    1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly   20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193  AGCCCCCACCAAGCCGCCGCCGCCGTCCGTAGACCAGACCCCCAGGACCCTGGCCACCATGGGC  252

Qy   21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla   40
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  253  CAGAGAGCATTACCTTCATCTCTGCTCTTGCTGCTCTCGAGTCGCGGCCCTTGAGTGCCCCACCTGCT  312
```

```
-------------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1         18058: contig of 18058 bp in length
*    18059         18158: gap of 100 bp
*    18159         35144: contig of 16986 bp in length
*    35145         35244: gap of 100 bp
*    35245         54710: contig of 19466 bp in length
*    54711         54810: gap of 100 bp
*    54811         72936: contig of 18126 bp in length
*    72937         73036: gap of 100 bp
*    73037         92888: contig of 19852 bp in length
*    92889         92988: gap of 100 bp
*    92989        108739: contig of 15751 bp in length
*   108740        108839: gap of 100 bp
*   108840        111738: contig of 2899 bp in length.
```

FEATURES	Location/Qualifiers
source	1..111738
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="1"
	/clone="RP3-426N7"
	/clone_lib="RPCI-3"
misc_feature	1..18058
	/note="assembly_fragment:02048
	fragment_chain:1
	clone_end:T7
	vector_side:left"
misc_feature	18159..35144
	/note="assembly_fragment:02454
	fragment_chain:1"
misc_feature	35245..54710
	/note="assembly_fragment:02786
	fragment_chain:1"
misc_feature	54811..72936
	/note="assembly_fragment:00223
	fragment_chain:2"
misc_feature	73037..92888
	/note="assembly_fragment:01820
	fragment_chain:2"
misc_feature	92989..108739
	/note="assembly_fragment:01122
	fragment_chain:2"
misc_feature	108840..111738
	/note="assembly_fragment:02919
	fragment_chain:2
	clone_end:SP6
	vector_side:right"

ORIGIN

Alignment Scores:
```
Pred. No.:                                    Length:           111738
Score:                 2.52e-24               Matches:          211
Percent Similarity:    1128.00                Conservative:     0
Best Local Similarity: 99.5%                  Mismatches:       1
Query Match:           99.5%                  Indels:           0
DB:                    99.5%                  Gaps:             0
                       14
```

US-09-989-890-238 (1-212) x AL592304 (1-111738)

```
Qy      1 SerProHisGlnAlaAlaAlaProValAlaProGlnThrProArgThrLeuAlaThrMetGly 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  95829 AGCCCCACCAAGCCGCGCACCCGACCCCGCTGAGACCCCAAGGACCCTGGCCACCATGGGC 95888

Qy     21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyValAlaGlyLeuProSerAlaAla 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCGGGCGGCCTTCCTTCCGCCCGCT 372

Qy     61 AlaGlyIleAlaSerArgAlaValAlaGluProValCysGlyAspAlaAlaProAlaCysLeu 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373 GCCGGGGATTGGCCTCCAGCGCGTGTGCGTGTCGCGGGGATGCAGCCCCTGCCTGTCTA 432

Qy     81 LeuArgThrProLeuArgGlyLeuLeuLysProThrHrGlyProArgSerThrMetGluCys 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433 CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGCCAAGGAGCACAATGAGTGC 492

Qy    101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPro 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    493 CCCCAGCCCTGATCGTGCACCCCCCAGCCGCGGCGGGATGCCAGCAGCTCAAGTCAACCA 552

Qy    121 TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLeuAlaSerLeuCysIleProThr 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    553 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC 612

Qy    141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle 160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    613 CGAGGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGCTCGCAAGGAGCCACTGGCCGATC 672

Qy    161 ProHisProCysAspThrArgAlaCysProAlaProLeuProValValLeuValAlaProArg 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    673 CCCACCCCATGCGACACAGCCTGCCCAGCACCCTTTGCCAGTAGTCCTCGTGGCTCCGAGG 732

Qy    181 SerThrIleLeuSerMetSerSerArgThrTrpThrCysSerArgArgTrpAlaValAlaProCys 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    733 AGTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGCAGTGGCTCCATGT 792

Qy    201 ArgAlaGlyLeuLysMetCysSerSerSerArgSer 212
            ||||||||||||||||||||||||||||||||||||
Db    793 CGAGCCGGAGAATTGATGTGCTCATCTTCAAGAAGC 828
```

```
RESULT 7
AL592304
LOCUS       AL592304              111738 bp    DNA     linear   HTG 25-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP3-426N7, 7 unordered pieces.
ACCESSION   AL592304
VERSION     AL592304.1  GI:14586390
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
AUTHORS     Mclay,K.
TITLE       Direct Submission
JOURNAL     Submitted (24-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ426N7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 110287 bases at least Q40
            Consensus quality: 110500 bases at least Q30
            Consensus quality: 110681 bases at least Q20
            Insert size: 111738; sum-of-contigs
            Insert size: 119403; 8.4% error; agarose-fp
            Quality coverage: 11.23x in Q20 bases; sum-of-contigs Quality
            coverage: 10.67x in Q20 bases; agarose-fp
```

```
Db  95889 CAGAGAGCATTACCTTCATCTCGGCTCTGCTGAGTCCCGCCGGCCCTGAGTCCCACCTGCT 95948
                                                                      Qy   1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
                                                                             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                      Db 357 AGCCCCACCAAGCCGCCGCACCCGCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 416

Qy  41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60    Qy  21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 95949 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCGTCTTCCGCCGCT 96008    Db 417 CAGAGAGCATTACCTTCATCTCGGCTCTGCTGAGTCCCGCCGGCCCCCACCTGCT 476

Qy  61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80    Qy  41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla-Al  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96009 GCCGGGATTGCCTCCAGCGTGTGGAGCCTGTGTGGGATGCAGCCCCTGCCTGTCTA 96068    Db 477 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCGTCTTCCGCCGC 536

Qy  81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100   Qy  60 aAlaGlyIleAlaSerSerAlaValGlu-ProValCysGlyAspAlaAlaProAlaCysL  80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||          ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db 96069 CTGAGGATTCCACTGAGGGGACTGCTGAAGCCAACTGGCCCAAGGAGCACAATGGAGTGC 96128    Db 537 TGCCGGGATTGCCTCCAGCGTGTGGAGCCTGTGTGGGATGCAGCCCCTGCCTGTC 596

Qy 101 ProProAlaLeuIleValHisProAlaGlyMetAlaSerGlySerGlnPro  120   Qy  80 euLeuArgThr-ProLeuArgGlyLeuLeuLysProThrGly-ProArgSerThrMetGl  99
        ||||||||||||||||||||||||||||||||||||||||||||||||||          ||||||||| |||||||||||||||||||||| |||||||||||||||||
Db 96129 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGCTCAAGTCAACCA 96188    Db 597 TACTGAGGATCCCCACTGAGGGGACTGCTGAAGCCAACTGGTCGCCAAGGAGCACAATGGA 656

Qy 121 TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140   Qy  99 uCysProProAlaLeuIleValHisProPro-AlaGlyMetAla---SerGlySerS  118
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||          ||||||||||||||||||||||||||||||||| ||||||||||   :::
Db 96189 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTACC 96248    Db 657 GTGCCCCCCAGCCCTGATCGTGCACCCCCCAGACCGGCGGCTCAAGTGGCCAGGCGGGCTGCAA 716

Qy 141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisProIle  160   Qy 118 erGlnProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysI  138
        ||||||||||||||||||||||||||||||||||||||||||||||||||||          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96249 CGAGGGGCCACCTCCCCAGCCCTGCGATGCGGAGCCACTGGCCGGATC 96308    Db 717 GTCAACCATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTA 776

Qy 161 ProHisProCysAspThrAlaProLeuProValValLeuValAlaProArg  180   Qy 138 leProThrArg-GlyProProGlnProLeuMetArgThrProAlaAlaArgSerHis  157
        ||||||||||||||||||||||||||||||||||||||||||||||||||          ||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db 96309 CCCCACCCCATGCGACACACCCTGCCCAGACCTTGCCGGTCCGAGG 96368    Db 777 TCCCTACCCGGAGGGGCCACCTCCCCAGCCCTGCGATGCGGGACTCCTGCTGCAAGGAGCCAC 836

Qy 181 SerThrIleLeuSerMetSerArgArgThrTrpThrCysArgArgTrpAlaValAlaProCys  200   Qy 158 TrpProIleProHisProCysAsp-ThrAlaCysProAlaProLeuProValLeuVa  177
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||          ||||||||||||||||||| :::|||||||||||||||||||||||||||
Db 96369 AGTACTATTCTTTCCATGAGTCGGACCTGGACCTGGTGGCTCCATGT 96428    Db 837 TGGCCGATCCCCACCCCATGCGACAGACCCGCCCAGCACCTTTGCCAGTAGTCCTCGT 896

Qy 201 ArgAlaGluLysLeuMetCysSerSerSerArgSer 212   Qy 177 lAlaProArgSerThrIleLeuSerMetSerArgArgThrTrpThrCysArgArgTrpAlaVa 197
        ||||||||||||||||||||||||||||||||||          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96429 CGAGCCGCCGGAAAATTGATGTGCTCATCTTCAAGAGAGC 96464    Db 897 GGCTCCGAGGAGTACTATTCTTTCCATGAGTCGGACCTGGACCTGGCCGGAGATGGGCAGT 956

                                                                      Qy 197 lAlaProCysArgAlaGluLysLeuMetCysSerSerSerArgSer 212
                                                                             ||||||||||||||||||||||||||||||||||||||||||||||
                                                                      Db 957 GGCTCCCATGTCGAGCCGCCGGAGAAATTGATGTGCTCATCTTCAAGAAGC 1002
```

RESULT 8
AX535029
LOCUS   AX535029   1977 bp DNA   linear   PAT 22-NOV-2002
DEFINITION Sequence 56 from Patent WO02068633.
ACCESSION AX535029
VERSION   AX535029.1 GI:25261664
KEYWORDS .
SOURCE   Homo sapiens (human)
 ORGANISM Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
        Hominidae; Homo.
REFERENCE 1
 AUTHORS   Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.
 TITLE   Compositions and methods relating to lung specific genes and
        proteins
 JOURNAL   Patent: WO 02068633-A 56 06-SEP-2002;
        Diadexus, Inc. (US)
FEATURES   Location/Qualifiers
 source   1..1977
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:   5.22e-22   Length: 1977
Score: 1002.50   Matches: 206
Percent Similarity: 94.5%   Conservative: 2
Best Local Similarity: 93.6%   Mismatches: 4
Query Match: 88.4%   Indels: 8
DB: Gaps: 1

US-09-989-890-238 (1-212) x AX535029 (1-1977)

RESULT 9
AX535090
LOCUS   AX535090   1977 bp DNA   linear   PAT 22-NOV-2002
DEFINITION Sequence 117 from Patent WO02068633.
ACCESSION AX535090
VERSION   AX535090.1 GI:25261789
KEYWORDS .
SOURCE   Homo sapiens (human)
 ORGANISM Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
        Hominidae; Homo.
REFERENCE 1
 AUTHORS   Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.
 TITLE   Compositions and methods relating to lung specific genes and
        proteins
 JOURNAL   Patent: WO 02068633-A 117 06-SEP-2002;
        Diadexus, Inc. (US)
FEATURES   Location/Qualifiers
 source   1..1977
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:   5.22e-22   Length: 1977
Score: 1002.50   Matches: 206
Percent Similarity: 94.5%   Conservative: 2

Best Local Similarity: 93.6%   Mismatches: 4
Query Match: 88.4%   Indels: 8
DB: Gaps: 1

US-09-989-890-238 (1-212) x AX535090 (1-1977)

```
Qy    1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  357 AGCCCCCACCAAGCCGCCGCACCCGTAGACCAGGACCCCAGGACCCTGGCCACCATGGGC 416

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuSerArgArgProLeuSerProSerProAla 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  417 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGGCCCCTTGAGTCCCCACCTGCT 476

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla-Al 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db  477 GCCTGCTCTGGCGACCCTGGGTGTGGAGTGGTGGCCGGGTGCCTCTTCTGCTTCCGCCGC 536

Qy   60 aAlaGlyIleAlaSerAlaValGlu-ProValCysGlyValAspAlaAlaProAlaCysL 80
         ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  537 TGCCGGGATTGCCTCCAGCGGTGTCTCCAGCCGTGTGTGCGGGGATGCAGCCCCTGCCTGTC 596

Qy   80 euLeuArgThr-ProLeuArgGlyLeuLeuLysProThrGly-ProArgSerThrMetGl 99
          ||||||||||| |||||||||||||||||||||||||||||| |||||||||||||||||
Db  597 TACTGAGGACTCCCACTGAGGGGACTGCTGAAGCCAACTGGTGCCAAGGAGCACAATGA 656

Qy   99 uCysProProAlaLeuIleValHisProPro-AlaGlyGlyMetAla---SerGlySerS 118
          |||||||||||||||||||||||||||||| ||||||||||||||    :::|
Db  657 GTGCCCCCCAGCCCTGATCGTGCACCCCCAGACCGGCGGGATGGCCAGGCGGGCTGCAA 716

Qy  118 erGlnProTrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysI 138
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  717 GTCAACCATGGGCACGGCAGCAGCTTCAGTGACTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTA 776

Qy  138 leProThrArg-GlyProProProGlnProLeuMetAlaArgThrProAlaAlaArgSerHis 157
          |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db  777 TCCCTACCCGAGAGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCAAGGAGCCAC 836

Qy  158 TrpProIleProHisProCysAsp-ThrAlaCysProAlaProLeuProValValLeuVa 177
          ||||||||||||||||||||||||| :::|||||||||||||||||||||||||||||||
Db  837 TGGCCGGATCCCCCACCCATGCGAGCAGCACACCCCAGCACCCTTTGCCAGTAGTCCTCGT 896

Qy  177 lAlaProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaVa 197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  897 GGCTCCGAGGAGTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGCAGT 956

Qy  197 lAlaProCysArgAlaGlyLysLeuMetCysSerSerSerArgSer 212
          |||||||||||||||||||||||||:::|||||||||||||||||||
Db  957 GGCTCCCATGTGCGAGCGGCGAGAAATTGATGTGCTCATCTTCAAGAGAGC 1002
```

RESULT 10
AL627228
LOCUS   AL627228   193813 bp   DNA   linear   ROD 04-FEB-2003
DEFINITION   Mouse DNA sequence from clone RP23-137L22 on chromosome 4, complete
             sequence.
ACCESSION   AL627228
VERSION   AL627228.31 GI:28208155
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 193813)
AUTHORS   Hopkins,B.
TITLE   Direct Submission
JOURNAL   Submitted (04-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquery@sanger.ac.uk clone requests: clonerequest@sanger.ac.uk
COMMENT   On Feb 4, 2003 this sequence version replaced gi:22474404.
             ---------- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was finished as follows unless otherwise noted: all regions were
either double-stranded or sequenced with an alternate chemistry or
covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at least one
plasmid subclone or more than one M13 subclone; and the assembly
was confirmed by restriction digest, except on the rare occasion of
the clone being a YAC.
RP23-137L22 is from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.

FEATURES   Location/Qualifiers
     source   1..193813
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="4"
                     /clone="RP23-137L22"
                     /clone_lib="RPCI-23"

ORIGIN

Alignment Scores:
Pred. No.:   Length: 193813
Score: 5.49e-11   Matches: 143
Percent Similarity: 685.00   Conservative: 8
Best Local Similarity: 71.2%   Mismatches: 61
Query Match: 67.5%   Indels: 0
DB: 60.4%   Gaps: 0
                        9

US-09-989-890-238 (1-212) x AL627228 (1-193813)

```
Qy       1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
             |||||||||:::||       |||||||||||||||||||||||||||||||||||||
Db  103992 AGCCCCCACCAAGCCGCCGTCGCACCAGGAGGCCAGACCCTAGGACCCTGGCCACCACGGGC 104051

Qy      21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40
             ||||||               |||||:::::|||||||||||||||||||||||||||
Db  104052 CAGAGAGTATCACCTTCATTCAGGTCTGCAGAACCAGCCAACGAGCCCCCAACCTGCT 104111

Qy      41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
             |||||:::||             |||||||::|||||||||||||||||||||||||||
Db  104112 GCCTCCTCTGGCGACCCTGGGGGTTGGGACTGGTGAGGCTGTGCCTTCTGCTTCCGACGCT 104171

Qy      61 AlaGlyIleAlaSerAlaValAlaGluProValCysGlyValAspAlaAlaProAlaCysLeu 80
             |||||||||||||||||||||       ||||||||||||||||||||||||||||||||
Db  104172 GCAGGGATTGCCTCGAGCCGGTGTCTGTGTGCGCGGGTTGTCAGCCCCTGCTTGCTCTG 104231

Qy      81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
             ||||||           |||||:::|||||||||||||:::||||||||||||||||||
Db  104232 CCGGAGACCCCATTGAAGGGTCTGCGAAGCCCTGCCTGGCCAAGGAACACAATGGTGTGC 104291

Qy     101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnInPro 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104292 CCCCAGCCCAGCCCCCCAGCCCCGGATGGCCGCGGGATGGCCAGAGGCTCAAGACAAGCA 104351
```

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104352 TGGGCAGCAGCTTCAGCTACCCCAGTGATGTTAAGCTCAAAGGCATCCCCTGTCTACCCCTACC 104411

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104412 GCCATGCCACCTCCCCAGTCCCTGCACGTGGACTCCTGCTGCCAAGGAGCCCCTGGCCGAGC 104471

Qy   161 ProHisProCysAspThrAlaCysProAlaProLeuProValLeuValAlaProArg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104472 CTCCTCCCACGGCACAGCTTGCCTAGCACCTTCACCAACAGCCCCGCGGCTCTGAGG 104531

Qy   181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys 200
           |||| |||||| |||||:::|||||||||||||||||||||||||||||||||||||
Db 104532 AGTACTACTCCTTCCATGAATCGGACCTGGACCTGCCTGGAGATGGCCAGTGGGCAGTGGCTCCATGT 104591

Qy   201 ArgAlaGluGlyLeuMetCysSerSerSerArgSer 212
           ||| :::|||||||||||||||||||||||||||||
Db 104592 CGAGCCGGGAGATCGACGTGCCTTATTTCAAGAAGC 104627

RESULT 11
BC006890
LOCUS   BC006890   1680 bp   mRNA   linear   ROD 29-JUN-2004
DEFINITION   Mus musculus RIKEN cDNA 1810019J16 gene, mRNA (cDNA clone MGC:11921
             IMAGE:3599314), complete cds.
ACCESSION   BC006890
VERSION   BC006890.1   GI:13905189
KEYWORDS   MGC.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1680)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
             Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
             Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
             Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
             Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
             Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
             Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
             Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
             Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
             McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
             Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
             Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
             Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
             Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
             Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
             Dickson,M.C., Rodriguez,A.C., Grinwood,J., Schmutz,J., Myers,R.M.,
             Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
             Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
             human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2 (bases 1 to 1680)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-APR-2001) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Baylor College of Medicine Human Genome
             Sequencing Center
             Center code: BCM-HGSC

Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne,P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 16 Row: p Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 19526881.

FEATURES   Location/Qualifiers
     source   1..1680
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:11921 IMAGE:3599314"
                     /tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
                     ductal carcinoma. 5 month old virgin mouse."
                     /clone_lib="NCI_CGAP_Mam6"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene   1..1680
                     /gene="1810019J16Rik"
                     /db_xref="GeneID:69073"
                     /db_xref="MGI:1916323"
     CDS   142..792
                     /gene="1810019J16Rik"
                     /codon_start=1
                     /product="RIKEN cDNA 1810019J16"
                     /protein_id="AAH06890.1"
                     /db_xref="GI:13905190"
                     /db_xref="GeneID:69073"
                     /db_xref="MGI:1916323"
                     /translation="MSAPSPHRAVAPGGQTLRTLATMGQRVSPSFQALQNQPTSPQPA
                     ASSGAPGVGTGVGLPSASAAGLACSAVGLVCGAAAPAYLPETPLKGLPKPPGPRNTM
                     VCPPARTVHPPAAGMARGSRPAWAAASATLMLSSKASLSTPTAMPPPQSLTWTPAARS
                     PWPSLLPHGTACLAPSPTAPAALRSTTPSMNRTWTCLRWAVAPCRAGRSTCLFSRS"

ORIGIN

Alignment Scores:
Pred. No.:   1.7e-12   Length: 1680
Score: 680.00   Matches: 143
Percent Similarity: 71.2%   Conservative: 8
Best Local Similarity: 67.5%   Mismatches: 61
Query Match: 60.0%   Indels: 0
DB: 9   Gaps: 0

US-09-989-890-238 (1-212) x BC006890 (1-1680)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
           |||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db   154 AGCCCCCACCCGGGCCGTCGCACCAGGAGGCCAGACCCTAAGGACCCTGGCCACCATGGGC 213

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerArgProProAla 40
           |||||||||||||||:::||||||||||||::::|||||||||||||||||||||||||||
Db   214 CAGAGAGAGTATCACCTTCATTTCAGGTCTTCAGAACCAGCCAACGAGCCCAACCAACCTGCT 273

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   274 GCCTCCTCCTGGCGGCCCTGGGGTTGGGACTGGTGGTGAGGGCTGGTGCGGGGCTGGTCGGGCGCT 333

Qy   61 AlaGlyIleAlaSerSerAlaValAlaGluProValCysGlyAlaProAlaProAlaCysLeu 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   334 GCAGGGATTGCCTCAGCCGTCTGTGGGGGCTGTGTGCGGGGCTGCAGCCCGTGCTTATCTG 393

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProGlyThrGlyProArgSerThrMetGluCys 100
           ||||||||||:::|||||:::||||||||||||::::|||||||||||||||||||||||||
Db   394 CCGGAGACCCCATTGAAGGTCTGCGCAAGGCCGCCTGGGCAAGGACAACAATGGTGTGC 453

Qy   101 ProAlaLeuIleValHisProProAlaGlyMetAlaSerGlySerGlnPro 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db    454  CCCCCAGCCCGGACCGTGCGACCCCCCAGCCGCGGATGCCAGAGGCTCAAGACCAGCA  513
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121  TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140

Db    514  TGGGCAGCAGCTTCAGCTACCCTGATGTTAAGCTCAAGAGCATCCTGTCTACCCCTACC  573
               |||||
Qy    141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProProIle  160

Db    574  GCCATGGCCACCTCCCCAGTCCCTGACGTGGACTCCTGCTGCAAGGAGCCCCTGGCCGAGC  633
               |||||||
Qy    161  ProHisProCysArgProAlaCysProAlaProLeuProValValLeuProAlaAlaProArg  180

Db    634  CTCCTCCACACGGCACAGCTTGCCTAGCACCTTCACCAACAGCCCCGCGGGTCTGAGG  693
               |||  ||||
Qy    181  SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys  200

Db    694  AGTACTACTCCTTCCATGAATCGGACCTGCCTGACCTGCGAGATGGGCAGTGGCTCCATGT  753
               |||||||||||||||||||||||||
Qy    201  ArgAlaGluLysLeuMetCysSerSerSerArgSer  212

Db    754  CGAGCCGGGAGATCGACGTGCTTATTTTCAAGAAGC  789
               :::  |
```

RESULT 12
BC079148
LOCUS   BC079148   1718 bp   mRNA   linear   ROD 15-SEP-2004
DEFINITION Rattus norvegicus similar to hypothetical protein FLJ34633, mRNA
            (cDNA clone MGC:94165 IMAGE:7128510), complete cds.
ACCESSION   BC079148
VERSION   BC079148.1 GI:50927712
KEYWORDS   MGC.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 1718)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED   12477932
REFERENCE   2 (bases 1 to 1718)
  AUTHORS   Director MGC Project.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-AUG-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Howard Jacobs
            cDNA Library Preparation: Express Genomics
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305

            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 184 Row: d Column: 16
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES   Location/Qualifiers
     source   1..1718
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="MGC:94165 IMAGE:7128510"
                     /tissue_type="Kidney, rat (Brown Norway)"
                     /clone_lib="NIH_MGC_236"
                     /lab_host="DH10B"
                     /note="Vector: pExpress1"
     gene   1..1718
                     /gene="MGC94165"
                     /db_xref="GeneID:313018"
     CDS   73..1266
                     /gene="MGC94165"
                     /codon_start=1
                     /product="similar to hypothetical protein FLJ34633"
                     /protein_id="AAH79148.1"
                     /db_xref="GI:50927713"
                     /db_xref="GeneID:313018"
                     /translation="MPRFGQPRPSSGPPRLGPWERPTELCLETNDERSQPPGRRTRR
                     PDPKDPGHHGPESITFISGSABPANEPPTCCLLWRPWGWDWCRAAFCFRRCRDCLQRC
                     GACVRSCSPCLSAGDPIEGSSEAWAKEHNGVPPSPDRAPPSRRDGQKLKTSMGSSFS
                     YPDVKLKGIPVYPYRHATSPVPDADSCCKEPLADPPTRHSLPSTFTSSPRGSEBYYS
                     FHESDLDLPEMGGSGSMSSRBIDVLIFKHQIDELAKCTSDTVFLEKTSKIS
                     DLISSITQDYHLDBQDAEGRLVRGIIRISTRKSRSRPQTSEGRSARSTAPAAAPDSGH
                     ETMVGSGLSQDEJTVQISQETTADAIARKLRPYGAPGYPASQDSSFQGTDTDSSGAPL
                     LQYVC"
ORIGIN

Alignment Scores:
Pred. No.:   1718
Score: 1.1e-11   Length: 1718
                    653.00   Matches: 138
Percent Similarity: 69.3%   Conservative: 9
Best Local Similarity: 65.1%   Mismatches: 65
Query Match: 57.6%   Indels: 0
DB: 9   Gaps: 0

US-09-989-890-238 (1-212) x BC079148 (1-1718)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly   20
             |||||||||||||||||||||||||||||||||
Db   176 AGCCCCCACCAGGCCGTCGCACCCGAGGCCGACCCGGGGGTCCAGACCCTGGCCCACCACGGC 235

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla   40
             |||||| ||:::|||
Db   236 CAGAGAGCATCACCTTCATTTCAGGCTCTGCAGAACCAGCCAACAGCGCCCCAACCTGCT 295

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla   60
             |||||:::|||
Db   296 GCCTGCTCCTGGCACCTGGGTGGAGGGTGTAGGGCTGCTTCTTCCGACGCT 355

Qy   61 AlaGlyIleAlaSerSerAlaValAlaGluProValCysGlyAspAlaProAlaCysLeu   80
             ||| ||||||||||||
Db   356 GCAGGGATTGCTCACGGCTTCCAGCCGCTGTGGAGCGTGTTGTGCGGGAGCTGAGTGCG 415

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
             ::|:|||
Db   416 CTTGGAGACCCCATTGAAGGGTCTTCGGAAGCCGCCTGGCCAAGAACAACAACGGTGTGC 475

Qy   101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro 120
             ||||||
```

```
Db   476  CTCCCAGCCCAGACCGGAGGACCCCCAGCCGCCGGATGGCAAAAGTTCAAGACCAGCA  535
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   121  TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140

Db   536  TGGGCAGCAGCTTCAGCTACCCGATGTTAAGCTCAAGGGCATTCCCGGTCTACCCCTACC  595
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160

Db   596  GCCATGGCCACCTCCCCCGTCCCTGACGCGGACTCCTGCTGCAAGGAGCCCCTGGCAGACC  655
               ||||||||||||||||||||||||||||||||||||||||||
Qy   161  ProHisProCysAspThrAlaCysSerProAlaProLeuProValValAlaProAlaArg  180

Db   656  CTCCTCCCACACGGCACAGCTTGCCTAGCACCTTCACCAGCACCCCCGTGGCTCTGAGG  715
               ||||    |||||||
Qy   181  SerThrIleSerMetSerArgThrTrpThrCysArgArgTrpCysAlaValAlaProCys  200

Db   716  AGTACTATTCCTTCCATGAATCGGACCTGCCCAAATGGGCAGTGGCTCCATGT  775
               ||:::    ||||||
Qy   201  ArgAlaGluLysMetCysSerSerSerArgSer  212
               ||||||
Db   776  CGAGCCGTGAGATCGACGTGCTTATTTTCAAGAGAGC  811
```

RESULT 13
AC095979

LOCUS	AC095979 235419 bp DNA linear HTG 09-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-11020, WORKING DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION	AC095979
VERSION	AC095979.7 GI:24818145
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE   1 (bases 1 to 235419)
AUTHORS   Muzny,D.Marie, Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,C., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,W., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,S.L., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,J., Li,I., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheswari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelemeh,O., Okwuonu,G., Oiarnpunsagoon,A., Pai,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,G., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,X., Dunn,D., von
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 235419)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 235419)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Nov 9, 2002 this sequence version replaced gi:22772470.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
--------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
--------- Project Information
Center project name: GBBU
Center clone name: CH230-11020
--------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 225536 bases at least Q40
Consensus quality: 227458 bases at least Q30
Consensus quality: 228649 bases at least Q20
Estimated insert size: 235013; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*   consists of 3 contigs. The true order of the pieces
*   is not known and their order in this sequence record is
*   arbitrary. Gaps between the contigs are represented as
*   runs of N, but the exact sizes of the gaps are unknown.
*   This record will be updated with the finished sequence
*   as soon as it is available and the accession number will
*   be preserved.
*   1   221492: contig of 221492 bp in length
*   221493 221592: gap of unknown length
*   221593 228651: contig of 7059 bp in length
*   228652 228751: gap of unknown length
*   228752 235419: contig of 6668 bp in length.
FEATURES   Location/Qualifiers
source   1..235419
                     /organism="Rattus norvegicus"

```
misc_feature        /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-110Z0"
                    2958. .3664
                    /note="clone_boundary
                    clone end:T7
                    site:EcoRI
                    end sequence:BH344795"
misc_feature        219911. .220573
                    /note="clone boundary
                    clone end:Sp6
                    site:EcoRI
                    end_sequence:BH344813"
gap                 221493. .221592
                    /estimated_length=unknown
gap                 228652. .228751
                    /estimated_length=unknown

ORIGIN
```

```
RESULT 14
AC118963/c
LOCUS       AC118963              249406 bp    DNA      linear    HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-180B4, WORKING DRAFT SEQUENCE.
ACCESSION   AC118963
VERSION     AC118963.4  GI:25009106
KEYWORDS    HTG; HTGS_PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 249406)
AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,G., Chen,H., Chen,Y., Chu,J.,
            Cleveland,C., Cockrell,R., Cox.C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P. Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangun,B., Mapua,P., Martin,K., Martin,R., Martinez,B.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,B., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
            Nwaokelemen,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Pfannkoch.C.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch.C.
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,J., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B. Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 249406)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (22-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 249406)
AUTHORS     Rat Genome Sequencing Consortium.
TITLE       Direct Submission
```

```
Alignment Scores:
Pred. No.:               5.72e-10          Length:          235419
Score:                   653.00            Matches:         138
Percent Similarity:      69.3%             Conservative:    9
Best Local Similarity:   65.1%             Mismatches:      65
Query Match:             57.6%             Indels:          0
DB:                      14                Gaps:            0
```

```
US-09-989-890-238 (1-212) x AC095979 (1-235419)

Qy    1 SerProHisGlnAlaAlaAlaProValAspProArgThrLeuAlaThrMetGly  20
         |||||||||| |||||  ||||||   |||||||||||||||||||||||||||
Db 83189 AGCCCCCCACCAGGCCGTGCACCCGGAGGCCAGACCCTAAGGACCCTGGCCACCACGGGC  83248

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla  40
         ||| |||||||||||    |||||  ||| :|| ||||||:||||||||||||||||||
Db 83249 CAGAGAGCATCACCTTCATTCAGCTCTGCAGAACCAACGAGCCCCAACCTGCT  83308

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
         ||| ||||| | |||: ||::||||::|||| ||||||||||||||||||||||||||||
Db 83309 GCCTCCTCTGGCGCCCCCTGGGGTTGGGACTGGTGTAGGGCTGCCTTTCTGCTTCCGACGCT  83368

Qy   61 AlaGlyIleAlaSerSerAlaValAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80
         ||| ||: ||| |||||||| |||||| |||||| |||| ||||||:||||| ||||| ||
Db 83369 GCAGGGATTGCCTCCAGCGCTGTGGAGCTTGTGTCGCGGAGCTGTAGCCCCTGCTTATCTG  83428

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
         ||| ||||||||:|| || :::||||| |||||||| ||||||||||||  :::||||||
Db 83429 CTGGAGACCCCATTGAAGGGTTCTTCGGAAGCCCGCCTGGGCCAAGGAACAACGGTGTGC  83488

Qy  101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPro  120
         |||| | ||: |  :||||||||||| |||||||| || |||||||||||||:||:||||
Db 83489 CTCCAGCCCAGACCGAGCACCCCCCAGCCGCCGCCGGATGGCAAAAGTCAAGACCAGCA  83548

Qy  121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
         |||||||:||| |||||||::|  :|| || ||| ||  |||| |||||||||||||
Db 83549 TGGGCAGCAGCTTCAGCTACCCTGATGTTAAGCTCAAGGGCATCCCGGTCTACCCCTACC  83608

Qy  141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
         ||||| ||  || ||  |   ||| |||||||| || | ||||||:||||||| ||||
Db 83609 GCCATGCCACCTCCCCCCGGCCCCTGACGCGGACTCCTGCTGCAAGGAGCCCCTGGCAGACC  83668

Qy  161 ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg  180
         ||||| ||||||:|||||| |||||||||  ||||||||||||||| |||||:|||||||
Db 83669 CTCTCCCACACGGCACAGCTTGCCTAGCACCTTCACCAGCAGCCCCCGTGGCTCTGAGG  83728

Qy  181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys  200
         ||| || ||:|| :::||:||:||:||  ||:|| ||| |:::|||||||||||:|||||
Db 83729 AGTACTATTCCTTCCATGAATCGGACCTGGACCCTGCCCGAAATGGGCAGTGGCTCCATGT  83788

Qy  201 ArgAlaGluLysLeuMetCysSerSerSerArgSer  212
         ||||| :: |||| ||||||||||||||||||||||
Db 83789 CGAGCCGTGAGATCGACGTGCTTATTTTCAAGAAGC  83824
```

```
Db   14533  GCCTCCTCCTGGCGCCCTGGGGTTGGGACTGGTGTAGGGCTGGTAGGGCTGCTTCGTGCTTCCGACGCT  14474
            |||   |||||                   ||||||     ||||||||||||||||||||||||  |||
Qy      61  AlaGlyIleAlaSerSerAlaValAlaGluProValCysGlyAspAlaAlaProAlaCysLeu          80
Db   14473  GCAGGGATGGCCTCCAGCGCTGTGAGCTGTGTGCGGAGCTGTAGCCCCTGCTTATCTG              14414
            ||||   ||||||||||||||||||| :::||||   |||||||  :::||||||||
Qy      81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys          100
Db   14413  CTGGAGACCCCATTGAAGGTCTTCGGAGCCGCCTGGGCCAAGGAACAACACGGTGTGC              14354
            |||  ||||||||||||::|||||||   ||||| |||||||||||||||||||
Qy     101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro          120
Db   14353  CTCCCAGCCCAGACCGAGCACCCCCAGCCCGCCGGGGATGGGCAAAAGTCAAGACCAGCA            14294
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121  TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr          140
Db   14293  TGGGCAGCAGCTTCAGCTACCCTGATGTTAAGCTCAAGGGCATCCCGGTCTACCCCTACC           14234
            |||||||||||||||||||||||||||||||::::|||||||||||||||||||||||||
Qy     141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle              160
Db   14233  GCCATGCACCTCCCCGGTCCTGACGCGGACTCCTGCTGCAAGGAGCCCCTGGCAGACC             14174
            ||||||     ||||||||||||||||||||||||||||||||||||||||||||||
Qy     161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg          180
Db   14173  CTCCTCCCACACGGCACAGCTTGCCTAGCACCTTCACCAGCACCCCCGTGGCTCTGAGG            14114
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181  SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys          200
Db   14113  AGTACTATTCCTTCCATGAATCGGACCTGACCTGCCGGGAAATGGGCAGTGGCTCCATGT           14054
            |||||||                   :::||||||  |||||||||||||||||||||
Qy     201  ArgAlaGluLysMetCysSerSerArgSer                                         212
            ||||||            :::           ||||||||
Db   14053  CGAGCCGTGAGATCGACGTGCTTATTTCAAGAAGC                                    14018
```

RESULT 15
AC126312/c
LOCUS   AC126312   256017 bp   DNA   linear   HTG 13-NOV-2002
DEFINITION   Rattus norvegicus clone CH230-1B9, *** SEQUENCING IN PROGRESS ***,
             3 unordered pieces.
ACCESSION   AC126312
VERSION   AC126312.4 GI:2941465
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 256017)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
             Barbaria,K., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
             Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
             Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
             Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
             Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
             Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
             Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
             Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
             Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
             Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
             Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
             Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
             Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
             Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
             Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
             Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.B.,
             Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
             Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,J.,
             Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
             Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
             Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
             Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,B.,

JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA

COMMENT   On Nov 15, 2002 this sequence version replaced gi:23269749.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          --------- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          --------- Project Information
          Center project name: GVHY
          Center clone name: CH230-180E4
          --------- Summary Statistics
          Assembly program: Phrap; version 0.990329
          Consensus quality: 213357 bases at least Q40
          Consensus quality: 215358 bases at least Q30
          Consensus quality: 217213 bases at least Q20
          Estimated insert size: 223109; sum-of-contigs estimation
          Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
          ----
          * NOTE: Estimated insert size may differ from sequence length.
          * NOTE: This is a 'working draft' sequence. It currently
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * consists of 1 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          * 249406: contig of 249406 bp in length.

FEATURES   Location/Qualifiers
     source   1..249406
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-180E4"
     misc_feature   1..1279
                     /note="wgs_contig"
ORIGIN

Alignment Scores:
Pred. No.:   5.99e-10
Score: 653.00   Length: 249406
Percent Similarity: 69.3%   Matches: 138
Best Local Similarity: 65.1%   Conservative: 9
Query Match: 57.6%   Mismatches: 65
DB: 14   Indels: 0
                                         Gaps: 0

US-09-989-890-238 (1-212) x AC118963 (1-249406)

```
Qy      1  SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly   20
           |||||||||||||||||||||||||    |||||||||   ||
Db  14653  AGCCCCCACCAGGCCGTCGCACCCGGACCGGAGCCCAGACCCCTAAGGAACCCTGGCCACCACGGGC  14594
Qy     21  GlnArgLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProLeuSerProProAla   40
           ||||||           ||||||||||||:::||||||||||||||||
Db  14593  CAGAGAGCATCACCTTCATTCAGGCTCTGCAGAACCAGCCAACGAGCCCCAACCTGCT         14534
Qy     41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla     60
```

```
                  AUTHORS   Massey,B., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
                            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
                            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
                            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
                            Oragunye,N., Oviedo.R., Pace,A., Payton,B., Peery,J., Perez,L.,
                            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
                            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
                            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
                            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Tang,H.,
                            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
                            Tansey,J., Taylor,C., Taylor,T., Telfrod,D., Thomas,N., Thomas,S.,
                            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
                            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
                            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
                            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
                            Weinstock,G. and Gibbs,R.
                  TITLE     Direct Submission
                  JOURNAL   Unpublished
                REFERENCE   2  (bases 1 to 256017)
                  AUTHORS   Worley,K.C.
                  TITLE     Direct Submission
                  JOURNAL   Submitted (05-JUL-2002) Human Genome Sequencing Center, Department
                            of Molecular and Human Genetics, Baylor College of Medicine, One
                            Baylor Plaza, Houston, TX 77030, USA
                REFERENCE   3  (bases 1 to 256017)
                  AUTHORS   Worley,K.C.
                  TITLE     Direct Submission
                  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
                            of Molecular and Human Genetics, Baylor College of Medicine, One
                            Baylor Plaza, Houston, TX 77030, USA
                  COMMENT   On Nov 13, 2002 this sequence version replaced gi:23267435.
                            The sequence in this assembly is a combination of BAC based reads
                            and whole genome shotgun sequencing reads assembled using Atlas
                            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
                            in the feature table below represents a scaffold (a 'contig-scaffold'
                            assembly (a 'contig-scaffold'). Within each contig-scaffold,
                            individual sequence contigs are ordered and oriented, and separated
                            by sized gaps filled with Ns to the estimated size. The sequence
                            may extend beyond the ends of the clone and there may be sequence
                            contigs within a contig-scaffold that consist entirely of whole
                            genome shotgun sequence reads. Both end sequences and whole genome
                            shotgun sequence only contigs will be indicated in the feature
                            table.
                            ----------- Genome Center
                            Center: Baylor College of Medicine
                            Center code: BCM
                            Web site: http://www.hgsc.bcm.tmc.edu/
                            Contact: hgsc-help@bcm.tmc.edu
                            ----------- Project Information
                            Center project name: TUNJ
                            Center clone name: CH230-1B9
                            ----------- Summary Statistics
                            Assembly program: Phrap; version 0.990329
                            Consensus quality: 214827 bases at least Q40
                            Consensus quality: 219116 bases at least Q30
                            Consensus quality: 221911 bases at least Q20
                            Estimated insert size: 222841; sum-of-contigs estimation
                            Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
                            ----------
                          * NOTE: Estimated insert size may differ from sequence length
                          *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
                          * NOTE: This is a 'working draft' sequence. It currently
                          *       consists of 3 contigs. The true order of the pieces
                          *       is not known and their order in this sequence record is
                          *       arbitrary. Gaps between the contigs are represented as
                          *       runs of N, but the exact sizes of the gaps are unknown.
                          * This record will be updated with the finished sequence
                          * as soon as it is available and the accession number will
                          * be preserved.
                          *      1   83721: contig of 83721 bp in length
                          *  83722   83821: gap of unknown length
                          *  83822  253567: contig of 169746 bp in length
                          * 253568  253667: gap of unknown length
```

```
FEATURES             Location/Qualifiers
     *   253668    256017: contig of 2350 bp in length.
     source          1..256017
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-1B9"
     misc_feature    6668..7475
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end sequence:BH274997"
     misc_feature    38358..39527
                     /note="wgs_contig"
     misc_feature    67950..69005
                     /note="wgs_contig"
     misc_feature    78781..81213
                     /note="wgs_contig"
     gap             83722..83821
                     /estimated_length=unknown
     misc_feature    83822..85574
                     /note="wgs_contig"
     misc_feature    219978..221353
                     /note="wgs_contig"
     gap             253568..253667
                     /estimated_length=unknown

ORIGIN
Alignment Scores:
Pred. No.:                                        256017
Score:                          6.12e-10   Length:              256017
Percent Similarity:             653.00     Matches:             138
Best Local Similarity:          69.3%      Conservative:        9
Query Match:                    65.1%      Mismatches:          65
DB:                             57.6%      Indels:              0
                                14         Gaps:                0

US-09-989-890-238 (1-212) x AC126312 (1-256017)

Qy     1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
         ||||||||||||||||||| || ||||||         |||||        ||
Db 155641 AGCCCCCACCAGGCCGTCGACCCCGGAGGCCAGACCCTAGGACCCTGGCCACCACGGGC 155582

Qy    21 GlnArgAlaLeuProSerSerLeuAlaLeuGlySerArgProLeuProProProAla  40
         |||||||||||||||| ||||  :::||||||||||     |||||||||
Db 155581 CAGAGAGCATCACCTTCATTCAGGTCTCGAGAACGCCCAACGAGCCCCCAACCTGCT 155522

Qy    41 AlaCysSerGlyAspProGlyCysGlyAlaGlyAlaGlyLeuProSerAlaSerAlaAla  60
         || ||||||  || |:|:||        |:::||       |||  ||||||
Db 155521 GCCTCCTCTGGCGGCCCCTGGGGTTGGGACTGGTGTAGGGCCTTCTGCCTTCCGACGCT 155462

Qy    61 AlaGlyIleAlaSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80
         ||||||||  |||:||||:||||||      ||:|||  ||||||:||||||
Db 155461 GCAGGGATTGCCTCCAGCGGCTGGAGCTTGTGTCGCGGAGCTGTAGCCCCTGCTTATCTG 155402

Qy    81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
         ||:::||||||    ||:||::|||      ||||:::||||||||
Db 155401 CTGGAGACCCCATTGAAGGGTCTTCGGAAGCGGCCTGGGGCCAAGGAACACACGGGTGTGC 155342

Qy   101 ProProAlaLeuIleValHisProProAlaAlaGlyGlyMetAlaSerGlySerGlnPro 120
         ||||:::||||||||      ||||||||   |||||  |||||||  |||||
Db 155341 CTCCCAGCCCAGACCGAGCACCCCCAGCCGCCGGGATGGGCAAAAGTCAAGACCAGCA 155282

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
         |||||| |:||||| |||  |::||||||||||||||:::||
Db 155281 TGGGCAGCAGCTTCAGCTACCTCAGCTATGATGTTAAGCTCAAGGGCATCCGGTCTACCCCTACC 155222

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIle 160
         |||  ||||:||||           |:||||||||||:|||||||
Db 155221 GCCATGCCACCTCCCCGTCCCTGACGGCGGGACTCCTGTCGCTCAAGGAGCCCCTGGCAACC 155162

Qy   161 ProHisProCysAspThrAlaCysSerProLeuProValValLeuValAlaProArg 180
         |||||||  ||| |:||||||||||| |||||||||||||||:|||||||
Db 155161 CTCCTCCCACACGGCACAGCTTGCCTAGCACCTTCACCAGCACCGCAGCCCCCGTGGCTCTGAGG 155102
```

```
Qy        181 SerThrIleLeuSerMetSerArgThrThrThrCysArgArgTrpAlaValAlaProCys 200
              |||||||||  |||||||:::||||||||||||    :::|||||||||||||||||||
Db     155101 AGTACTATTCCTTCCATGAATCGGACCTGGACCTGCCCGAAATGGGCAGTGGCTCCATGT 155042

Qy        201 ArgAlaGluLysLeuMetCysSerSerArgSer 212
              ||||||| :::     ||| ||||||||
Db     155041 CGAGCCGTGAGATCGACGTGCTTATTTTCAAGAAGC 155006

Search completed: March 17, 2006, 06:11:49
Job time : 7588 secs
```

```
            GenCore version 5.1.7
        Copyright (c) 1993 - 2006  Bioacceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      March 17, 2006, 03:48:50 ; Search time 588 Seconds
                              (without alignments)
                              2402.916 Million cell updates/sec

Title:          US-09-989-890-238
Perfect score:  1134
Sequence:       1 SPHQAAAPVDQTPRTLATMG........RRWAVAPCRAEKLMCSSSRS 212

Scoring table:  BLOSUM62
        Xgapop   10.0 , Xgapext   0.5
        Ygapop   10.0 , Ygapext   0.5
        Fgapop    6.0 , Fgapext   7.0
        Delop     6.0 , Delext    7.0

Searched:       4996997 seqs, 333346308 residues

Total number of hits satisfying chosen parameters:         9993994

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095245_16556/app_query.fasta_1
-DB=N_Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss04
-USER=US09989890 @CGN 1_1_727 @runat 16032006_095245_16556 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database :   N_Geneseq_21:*
```
            1:  geneseqn1980s:*
            2:  geneseqn1990s:*
            3:  geneseqn2000s:*
            4:  geneseqn2001as:*
            5:  geneseqn2001bs:*
            6:  geneseqn2002as:*
            7:  geneseqn2002bs:*
            8:  geneseqn2003as:*
            9:  geneseqn2003bs:*
           10:  geneseqn2003cs:*
           11:  geneseqn2003ds:*
           12:  geneseqn2004as:*
           13:  geneseqn2004bs:*
           14:  geneseqn2005s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	ID	Description
1	1134	100.0	791	14	AEA19545	Aea19545 Novel hum
2	1134	100.0	1713	8	ABT13390	Abt13390 Breast sp
3	1134	100.0	2392	10	ADD01260	Add01260 Human nuc
4	1128	99.5	1785	10	ADB62645	Adb62645 Human cDN

5	1065.5	94.0	654	6	ABT07645	Abt07645 Human bre
6	1002.5	88.4	1977	6	ABX92075	Abx92075 Lung spec
7	1002.5	88.4	1977	6	ABX92014	Abx92014 Lung spec
8	417	36.8	427	5	AAF66376	Aaf66376 Novel hum
9	c	178	15.7	1716	8	ACL72571	Acl72571 M. xanthu
10	c	178	15.5	9198	14	ACL64503	Acl64503 M. xanthu
11	173	15.3	2914	10	ADJ92146	Adj92146 Human hai
12	c	169	14.9	864	1	AAN60488	Aan60488 Eimeria t
13	c	169	14.9	864	2	AAQ03317	Aaq03317 cDNA of E
14	c	169	14.9	6567	2	AAQ03324	Aaq03324 Eimeria t
15	c	169	14.9	6567	8	ABZ22978	Abz22978 Eimeria m
16	c	169	14.9	6567	10	ABZ59095	Abz59095 E. maxima
17	c	168.5	14.9	2703	14	ACL66476	Acl66476 M. xanthu
18	c	168.5	14.9	23417	14	ACL64744	Acl64744 M. xanthu
19	168	14.8	1761	13	ADR07771	Adr07771 Full leng
20	c	166	14.6	1799	11	ADM03200	Adm03200 Human cDN
21	c	166	14.6	4262	5	AAF98697	Aaf98697 Human ova
22	c	166	14.6	5142	6	ABS76388	Abs76388 cDNA enco
23	c	166	14.6	5142	14	AEB35261	Aeb35261 Human Gef
24	166	14.6	22533	13	ADQ91713	Adq91713 Polyketid
25	166	14.6	22533	13	AEB87005	Aeb87005 Streptomy
26	166	14.6	22533	13	AEB86802	Aeb86802 Streptomy
27	c	166	14.6	71989	3	AAA29349	Aaa29349 Sorangium
28	166	14.6	164051	13	ADQ91695	Adq91695 Polyketid
29	166	14.6	164051	13	AEB86862	Aeb86862 Streptomy
30	166	14.6	164051	13	AEB86659	Aeb86659 Streptomy
31	c	165	14.6	3913	4	AAK52397	Aak52397 Human pol
32	c	165	14.6	68750	3	AAZ55887	Aaz55887 Sorangium
33	c	164	14.5	169144	10	ADL13748	Adl13748 Osteoarth
34	c	164	14.5	270150	11	ADP65796	Adp65796 Human 16p
35	163	14.4	20015	2	ABA20740	Aba20740 Human ner
36	c	162.5	14.3	15231	6	ABK63598	Abk63598 Rat seque
37	c	162.5	14.3	15231	10	ADB59205	Adb59205 Toxicity-
38	c	162.5	14.3	15231	10	ADB53026	Adb53026 Primary r
39	c	162.5	14.3	15231	10	ABT42078	Abt42078 Toxicity
40	c	162.5	14.3	15231	12	ADP72968	Adp72968 Renal tox
41	c	162.5	14.3	65140	4	AAD17184	Aad17184 Streptomy
42	162.5	14.3	125401	4	AAD17186	Aad17186 Streptomy
43	c	162	14.3	24120	8	ABX11642	Abx11642 Human ser
44	c	161.5	14.2	14800	8	ABL66291	Abl66291 Lung canc
45	c	161.5	14.2	14800	12	ADP13447	Adp13447 Renal cel

ALIGNMENTS

```
RESULT 1
AEA19545
ID   AEA19545 standard; cDNA; 791 BP.
XX
AC   AEA19545;
XX
DT   11-AUG-2005  (first entry)
XX
DE   Novel human polynucleotide SEQ ID NO 239.
XX
KW   vulnerary; CNS-gen.; gene therapy; diagnostic; forensic; mapping;
KW   DNA purification; protein purification; osteoarthritic; antiarthritic;
KW   osteopathic; musculoskeletal disease; osteoporosis; endocrine disease;
KW   periodontal disease; antiinflammatory; mouth disease; burns; injury;
KW   peripheral neuropathy; Alzheimers disease; neuroprotective; nootropic;
KW   degeneration; parkinsons disease; antiparkinsonian; neurological disease;
KW   cerebrovascular ischemia; cerebroprotective; vasotropic;
KW   cardiovascular disease; autoimmune disease; immunosuppressive;
KW   immune disorder; viral infection; virucide; infection; cancer;
KW   cytostatic; neoplasm; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200504806-A2.
XX
PD   02-JUN-2005.
XX
PF   11-MAR-2004; 2004WO-US007412.
```

```
Db   223  GCCTGCTCTCGGCGACCCTGGTGTGGGAGTGGTGCCGGCGGCTGCCTTCTGCTTCCGCCGCT  282

Qy    61  AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu     80
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   283  GCCGGGATTGCCTCCAGCGCTGTGGAGCCTGTGTGCGGGGATGCAGCCCCTGCCTGTCTA   342

Qy    81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys    100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   343  CTGAGGACTCCACTGAGGGGACTGCTGAAGCCCACTGGGCCAAGGAGCACAATGGAGTGC   402

Qy   101  ProProAlaLeuHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPro         120
               ||||||||||||||||||||||||||||||||||||||||||||
Db   403  CCCCCAGCCCTGCATCCTGCACCCCCCAGCGCCGGGGATGGCCAGCGGCTCAAGTCAACCA   462

Qy   121  TrpAlaAlaSerAlaThrProMetLeuSerLysAlaSerLeuCysIleProThr         140
               ||||||||||||||||||||||||||||||||||||||||||||
Db   463  TGGGCAGCAGCTTCAGCTACACCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC   522

Qy   141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle      160
               |||||||||||||||||||||||||||||||||||||||||||||
Db   523  CGAGGGCCACCTCCCCAGCCCTGATGCGGACCCCTGCTGCAAGAGCCACTGGGCCGATC   582

Qy   161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg   180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   583  CCCACCCATGGCACCACAGCCTGCCCAGCACCTTTGCCAGTAGTCCTCGTGGCTCCGAGG   642

Qy   181  SerThrIleLeuSerMetSerArgThrTrpProThrCysArgArgTrpAlaValAlaProCys   200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   643  AGTACTATTCTTCCATGAGTCGGACCTGGCCGACCTGCCGGAGATGGGCAGTGGCTCCATGT   702

Qy   201  ArgAlaGluLysMetCysSerSerSerArgSer                             212
               ||||||||||||||||||||||||||||||
Db   703  CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAGC                           738
```

RESULT 2
ABT13390
ID   ABT13390 standard; DNA; 1713 BP.
XX
AC   ABT13390;
XX
DT   30-JAN-2003 (first entry)
XX
DE   Breast specific related polynucleotide SEQ ID No 105.
XX
KW   Cytostatic; BSP-agonist; BSP-antagonist; vaccine; gene therapy; cancer;
KW   metastatic; breast cancer; breast specific; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277232-A2.
XX
PD   03-OCT-2002.
XX
PF   21-NOV-2001; 2001WO-US043815.
XX
PR   22-NOV-2000; 2000US-0252509P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S, Macina RA, Recipon H, Pluta J, Sun Y, Liu C;
XX
DR   WPI; 2003-018927/01.
XX
PT   New isolated nucleic acid molecule, useful for treating breast cancer,
PT   and diagnosing or monitoring the presence of metastases of breast cancer
PT   in a patient.
XX
PS   Claim 1; Page 250-251; 377pp; English.
XX
CC   The invention relates to a novel isolated nucleic acid molecule
CC   comprising: a sequence encoding a sequence comprising 11-1518 amino acids
CC   ; a sequence comprising 190-8144 bp; or a sequence that selectively
CC   hybridises to, or having at least 60% identity with the 11-1518 amino

XX
CC   The invention describes a new isolated polynucleotide (I) encoding a
CC   polypeptide with biological activity comprising: a nucleotide sequence of
CC   SEQ ID NOS: 1-567 (fully defined); a nucleotide sequence that hybridizes
CC   to the sequence of (I) under stringent hybridization conditions; or a
CC   nucleotide sequence having greater than 99% sequence identity with the
CC   sequence of (i). Also described are: a(n) (expression)vector comprising
CC   (I); a host cell genetically engineered to comprise (I) operatively,
CC   associated with a regulatory sequence that modulates expression of the
CC   polynucleotide in the host cell; an isolated polypeptide comprising a
CC   sequence of SEQ ID NOS: 568-1134 (fully defined), where the polypeptide
CC   is: a polypeptide encoded by (I); or a polypeptide encoded by a
CC   polynucleotide hybridizing under stringent conditions with any one of SEQ
CC   ID NOS: 1-567; a composition comprising the polypeptide of (3) and a
CC   carrier; an antibody directed against the polypeptide of (3); a method
CC   for detecting (I) in a sample; a method for detecting the polypeptide of
CC   (3) in a sample; a method for identifying a compound that binds to the
CC   polypeptide of (3); a method of producing the polypeptide of (3); and a
CC   collection of polynucleotides, where the collection comprising of at
CC   least one of SEQ ID NOS: 1-567. (I) is a polynucleotide comprising any of
CC   the sequences of SEQ ID NOS: 1-567 encoding a polypeptide with biological
CC   activity, which comprises any of the amino acid sequence of SEQ ID NOS:
CC   568-1134. All sequences are fully defined in the specification. The
CC   sequences and methods are useful in diagnostics, forensic, and gene
CC   mapping, in identifying of mutations responsible for genetic disorders or
CC   other traits, in assessing biodiversity, and for producing many other
CC   types of data and products dependent on DNA and amino acid sequences. The
CC   composition and method are useful for treating a disease or disorder,
CC   e.g. osteoporosis, osteoarthritis, periodontal disease, burns, CNS and
CC   peripheral disease, Alzheimer's disease, Parkinson's disease, stroke,
CC   autoimmune disorders, viral infection, or cancer. This sequence encodes a
CC   novel polypeptide of the invention.
XX
SQ   Sequence 791 BP; 149 A; 278 C; 222 G; 142 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 791
Score: 4.49e-36   Matches: 212
Percent Similarity: 1134.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 14   Gaps: 0

US-09-989-890-238 (1-212) x AEA19545 (1-791)

```
Qy     1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly    20
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   103  AGCCCCCACCAGGCGGCCGCACCCGTAGACCCAGACCCCAAGGACCCTGGCCACCATGGGC  162

Qy    21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla    40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   163  CAGAGAGCATTACCTTCATCTCTGCTCTCTGGCTCTCTGCTGAGCCGCCCTTGAGTCCCCACCTGCT  222

Qy    41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla    60
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
```

CC   acid or 190-8144 nt sequence. The polypeptide and the nucleic acid are
CC   useful for treating breast cancer, and diagnosing or monitoring the
CC   presence of metastases of breast cancer in a patient. The polynucleotides
CC   of the invention can be used to treat disorders by gene therapy. This
CC   polynucleotide represents a breast specific related sequence of the
CC   invention
XX
SQ   Sequence 1713 BP; 344 A; 553 C; 468 G; 348 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   8.54e-36
Score: 1134.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 100.0%
DB: 8

Length: 1713
Matches: 212
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x ABT13390 (1-1713)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   104 AGCCCCCACCAAGCCGCCCGCACCCGTAGACCAGGACCCCAAGGACCCTGGCCACCATGGGC 163

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla   40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   164 CAGAGAGCATTACCTTACCTTCTCGCTCTGCTGAGCCGGCCCCTTGAGTCCCCCACCTGCT 223

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCCGCT 283

Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu   80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   284 GCCGGGATTGCCTCCAGCCGCGTGTGAGCCTGTGTGCGGGATGCAGCCCCCTGCCTGTCTA 343

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   344 CTGAGGACTCCACTGAGGGACTGCTGAAGCCAACTGGCCAAGGAGCACAATGGAGTGC 403

Qy   101 ProProAlaLeuIleValHisProProAlaGlyIleGlyMetAlaSerGlySerGlnPro 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGCGGGATGGCCAGCGGGGCTCAAGTCAACCA 463

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLeuAlaSerLeuCysIleProThr 140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   464 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCGTGTATCCCTACC 523

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle 160
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   524 CGAGGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTGCCAAGAGCCACTGGCCGATC 583

Qy   161 ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   584 CCCACCCCATGCGACACCAGCCTGCCCAGCACCTTTGCCACTAGTCCTCGTGGCTCCGAGG 643

Qy   181 SerThrIleLeuSerMetSerArgThrThrTrpThrCysSerArgArgTrpAlaAlaProCys 200
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   644 AGTACTATTCTTTCCATGAGTCGGACCTGACCTGCCGGAGATGGGGAGTGGCTCCCATGT 703

Qy   201 ArgAlaGluLysLeuMetCysSerSerSerArgSer 212
             ||||||||||||||||||||||||||||||||||||
Db   704 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 739

RESULT 3
ADD01260
ID   ADD01260 standard; cDNA; 2392 BP.
XX
AC   ADD01260;
XX
DT   01-JAN-2004 (first entry)
XX
DE   Human nucleic acid-associated protein NAAP-41 cDNA SEQ ID NO:98.
XX

KW   human; nucleic acid-associated protein; NAAP; cytostatic;
KW   antiarteriosclerotic; anti-HIV; antiallergic; cerebroprotective;
KW   antiparkinsonian; anticonvulsant; nootropic; neuroprotective;
KW   antiinflammatory; ophthalmological; thyromimetic; antiarthritic;
KW   hepatotropic; antibacterial; virucide; protozoacide; antiparasitic;
KW   fungicide; gene therapy; cell proliferative disease; cancer;
KW   atherosclerosis; hepatitis; neurological disorder; Parkinson's disease;
KW   Alzheimer's disease; stroke; epilepsy; developmental disorder;
KW   renal tubular acidosis; anaemia; glaucoma; hypothyroidism;
KW   autoimmune disorder; inflammatory disorder; AIDS; allergy;
KW   atopic dermatitis; arthritis; infection; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200305421219-A2.
XX
PD   03-JUL-2003.
XX
PF   18-DEC-2002; 2002WO-US041115.
XX
PR   19-DEC-2001; 2001US-0343004P.
PR   11-JAN-2002; 2002US-0347633P.
PR   25-JAN-2002; 2002US-0351749P.
PR   22-FEB-2002; 2002US-0359498P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Azinzai Y,   Baughn MR,   Becha SD,   Borowsky ML,   Burford N;
PI   Elliott VS,   Emerling BM, Forsythe IJ, Gorvad AE,   Griffin JA;
PI   Kable AE,   Khare R,   Lal PG,   Lee EA,   Lee SY,   Li JX,   Marquis JP;
PI   Ramkumar J,   Richardson TW,   Sprague WW,   Swarnakar A,   Tang YT;
PI   Chawla NK,   Warren BA,   Yue H;
XX
DR   WPI; 2003-559157/52.
DR   P-PSDB; ADD01203.
XX
PT   New human nucleic acid-associated proteins (NAAP), useful for diagnosing,
PT   treating and preventing diseases or conditions associated with the
PT   aberrant NAAP expression, e.g. cancer, hepatitis, AIDS, atherosclerosis,
PT   infections.
XX
PS   Claim 5; SEQ ID NO 98; 405pp; English.
XX
CC   The present invention describes human nucleic acid-associated proteins
CC   designated NAAP-1 to NAAP-57. The human NAAPs have cytostatic,
CC   antiarteriosclerotic, anti-HIV, antiallergic, cerebroprotective,
CC   antiparkinsonian, anticonvulsant, nootropic, neuroprotective,
CC   antiinflammatory, ophthalmological, thyromimetic, antiarthritic,
CC   hepatotropic, antibacterial, virucide, protozoacide, antiparasitic and
CC   fungicide activities, and can be used in gene therapy. The NAAP protein
CC   and polynucleotide sequences can be used in diagnosing, treating and
CC   preventing diseases or conditions associated with the decreased
CC   expression or overexpression of NAAP, such as cell proliferative diseases
CC   (e.g. cancer, atherosclerosis, hepatitis), neurological disorders
CC   (Parkinson's disease, Alzheimer's disease, stroke, epilepsy),
CC   developmental disorders (renal tubular acidosis, anaemia, glaucoma,
CC   hypothyroidism), autoimmune/inflammatory disorders (AIDS, allergies,
CC   atopic dermatitis, arthritis) and infections (e.g. bacterial, viral,
CC   parasitic, protozoal, fungal). The present sequence encodes human NAAP-
CC   41, from the present invention.
XX
SQ   Sequence 2392 BP; 471 A; 743 C; 674 G; 504 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   1.13e-35
Score: 1134.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 100.0%
DB: 10

Length: 2392
Matches: 212
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x ADD01260 (1-2392)

```
PR    25-JAN-2002; 2002US-00350978.
XX
PA    (HELI-) HELIX RES INST.
PA    (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI    Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI    Yamamoto J, Isono Y, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI    Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR    WPI; 2003-450961/43.
DR    P-PSDB; ADB64615.
XX
PT    New polynucleotides and polypeptides, useful for developing a diagnostic
PT    marker or medicines for regulation of their expression and activity, or
PT    as targets of gene therapy.
XX
PS    Claim 1; Page 1; 222pp; English.
XX
CC    The invention discloses a polynucleotide comprising a sequence selected
CC    from 1970 fully defined nucleotide sequences which encode novel
CC    polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC    or its partial peptide, an antibody binding to the polypeptide or peptide
CC    of the polynucleotide, immunologically assaying the polypeptide or
CC    peptide of the polynucleotide by contacting the polypeptide or peptide
CC    with the antibody of the encoded protein, and observing the binding
CC    between the two, a transformant carrying the polynucleotide in an
CC    expressible manner and an antisense polynucleotide. The oligonucleotide
CC    is useful as a primer for synthesising the polynucleotide, or as a probe
CC    for detecting the polynucleotide. The polynucleotides and encoded
CC    proteins are useful as pharmaceutical agents and many disease-related
CC    genes may be included in them, for developing a diagnostic marker or
CC    medicines for regulation of their expression and activity, or as targets
CC    of gene therapy. The genes are involved in tissue and/or cell
CC    regeneration. Membrane proteins, signal transduction-related proteins,
CC    transcription-related proteins, disease-related proteins and genes
CC    encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC    neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC    the activity or expression of the encoded protein to treat diseases. The
CC    sequence presented is a cDNA of the invention. Note: Some of the sequence
CC    data for this patent is not represented in the printed specification, but
CC    is based on sequence information supplied by the European Patent Office.
XX
SQ    Sequence 1785 BP; 358 A; 567 C; 504 G; 356 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                        Length:          1785
Score:            1.51e-35        Matches:         211
Percent Similarity:  1128.00      Conservative:    0
Best Local Similarity: 99.5%      Mismatches:      1
Query Match:      99.5%           Indels:          0
DB:                               Gaps:            0

US-09-989-890-238 (1-212) x ADB62645 (1-1785)
```

```
Qy     1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   358  AGCCCCACCAAGCCGCGCCACCCGTAGACCAGACCCCAAGGACACCCTGGCCACCATGGGC  417

Qy    21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerArgProProAla  40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   418  CAGAGAGCAGCATTACCTTCATCTCTGGCTCTGCTCTTGAGTCCCCACCTGCT  477

Qy    41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   478  GCCTGCTCTGGCGACCCTGGGTGTGGAGGATGGTCCCGGCGTCCGCGTTCCGCCGCT  537

Qy    61  AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   538  GCCGGGATTGCCTCCAGCGTTGGAGCCTGTGCGGGGATGCAGCCCCTGCCTGCCTGTCTA  597

Qy    81  LeuArgThrProLeuArgGlyLeuLeuLysProThrArgLeuProArgSerThrMetGluCys  100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   598  CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCCAAGGAGCACAATGGGAGTGC  657

Qy   101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlyGlySerGlnPro  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   658  CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAGCGGCGGATCAACCA  717

Qy   121  TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   718  TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTATCCCTACC  777

Qy   141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIle  160
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   778  CGAGGGGCACCTCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGATC  837

Qy   161  ProHisProCysAspThrAlaCysProLeuProValValLeuValAlaProAlaArg  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   838  CCCCACCCATGCGACACAGCCTGCCCAGCACCTTGCCAGTAGTCCTGGGCTCCGAGG  897

Qy   181  SerThrIleLeuSerMetSerArgThrThrCysArgArgTrpAlaValAlaProCys  200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   898  AGTACTATTCTTTCCATGAGTCGGACCCTGGCACCTGCCGGAGATGGGCAGTGGCTCCATGT  957

Qy   201  ArgAlaGluLysLeuMetCysSerSerSerArgSer  212
             ||||||||||||||||||||||||||||||||||
Db   958  CGAGCCCGAGAAATTGATGTGCTCATCTTCAAGAAGC  993
```

```
RESULT 4
ADB62645
ID    ADB62645 standard; cDNA; 1785 BP.
XX
AC    ADB62645;
XX
DT    04-DEC-2003 (first entry)
XX
DE    Human cDNA encoding clone KIDNE20157100.
XX
KW    Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW    tissue regeneration; cell regeneration; membrane protein;
KW    signal transduction-related protein; transcription-related protein;
KW    osteoporosis; neurological disease; cancer; tumour.
XX
OS    Homo sapiens.
XX
FH    Key             Location/Qualifiers
FT    CDS             90..1286
FT                    /*tag= a
FT                    /product= "Clone KIDNE20157100 protein"
XX
PN    EP1308459-A2.
XX
PD    07-MAY-2003.
XX
PF    28-MAR-2002; 2002EP-00007401.
XX
PR    05-NOV-2001; 2001JP-00379298.
```

```
Qy   101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro 120
         ||||||||||||||||||||||||||||||              ||||||||||||
Db   493 CCCCAGCCCTGATCGTGCACCCCCCAGCCCGGCGGGATGGCCAGCCCAAGTCAACCA 552

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   553 TGGGCAGCAGCCTCAGCTACCCCGATGTTAAGCTTAAGCTCCTGTGTATCCCTACC 612

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIle 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   613 CGAGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTGCGCAAGGAGCCACTGGCCGATC 672

Qy   161 ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaValAlaProArg 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   673 CCCACCATGGCGACACAGCCTGCCCAGCACCTTTGCCAGTAGTTCCTCGTGGCTCCGGAGG 732

Qy   181 SerThrLeuSerMetSerArgArgThrTrpThrCysArgArgTrpAlaValAlaProCys 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   733 AGTACTATTCTTTCCATGAGTCGGACCTGCGACCTGCCGGAGATGGGCAGTGGCTCCATGT 792

Qy   201 ArgAlaGluLysLeuMetCysSerSerSerArgSer 212
         ||||||||||||||||||||||||||||||||||||
Db   793 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 828
```

RESULT 5
ABT07645
ID   ABT07645 standard; cDNA; 654 BP.
XX
AC   ABT07645;
XX
DT   14-NOV-2002 (first entry)
XX
DE   Human breast cancer associated coding sequence SEQ ID NO: 124.
XX
KW   Human; breast specific gene; breast specific protein; breast cancer;
KW   gene therapy; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20264611-A1.
XX
PD   22-AUG-2002.
XX
PF   12-FEB-2002; 2002WO-US004197.
XX
PR   13-FEB-2001; 2001US-0268292P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
PI   Sun Y, Liu C;
XX
DR   WPI; 2002-657582/70.
XX
PT   New breast specific nucleic acids and proteins, useful for identifying,
PT   diagnosing, monitoring, staging, imaging, and treating breast cancer and
PT   non-cancerous disease states in breast tissue, and in gene therapy.
XX
PS   Claim 1; Page 248; 367pp; English.
XX
CC   The present invention provides human breast specific coding sequences and
CC   proteins. These can be used in the diagnosis and treatment of breast
CC   cancer and non-cancerous diseases of the breast. The present sequence is
CC   a coding sequence of the invention
XX
SQ   Sequence 654 BP; 118 A; 232 C; 186 G; 118 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   1.71e-33   Length: 654
Score: 1065.50   Matches: 203
Percent Similarity: 95.8%   Conservative: 1
Best Local Similarity: 95.3%   Mismatches: 8

Query Match: 94.0%   Indels: 1
DB: 6   Gaps: 1

US-09-989-890-238 (1-212) x ABT07645 (1-654)

```
Qy     1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16 AGCCCCCACCAAGCGCCGCACCCGTGACCCCAGGACCCTGGCCACCATGGGC 75

Qy    21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    76 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGGCCCTTGAGTCCCCACCTGCT 135

Qy    41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   136 GCCTGCTCTGGCGACCCTGGTGTGGGAGTGGTGCCGGGGCGTGCCTTCTGCTTCCGCGCT 195

Qy    61 AlaGlyIleAlaSerAlaValAlaGluProValCysGlyValCysGlyValAspAlaAlaProAlaCysLeu 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   196 GCCGGGATTGCCTCCAGCGTGTGGGAGGGCGTGGCGGGGCAGCCCCTGCCTGTCTA 255

Qy    81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProThrGlySerThrMetGluCys 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   256 CTGAGGACTCCCCTGAGGGGACTGCTGAAGCAACTGGTCAAGAGCACAATGGAGTGC 315

Qy   101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAla---SerGlySerSerGln 119
         ||||||||||||||||              ||||              |||||::::|||
Db   316 CCCCAGCCCTGATCGTGCACCCCCCAGCCCGGCGGGATGGCCAGCCCGGCTGCAAGTCAA 375

Qy   120 ProTrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIlePro 139
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   376 CCATGGCAGCAGCCTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTATCCCT 435

Qy   140 ThrArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpPro 159
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436 ACCGAGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTGCGCAAGGAGCCACTGGCCG 495

Qy   160 IleProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaValAlaPro 179
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   496 ATCCCCACCCCACCAGCGGAGCACAGCCTGCCCAGCACCTTTGCCAGTAGTTCCTCGTGGCTCCG 555

Qy   180 ArgSerThrIleLeuSerMetSerArgArgThrTrpThrCysArgArgTrpAlaValAlaPro 199
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   556 AGGAGTACTATTCTTTCCATGAGTCGGACCTGACCTGCCCGAGATGGGCAGTGGCTCCA 615

Qy   200 CysArgAlaGluLysLeuMetCysSerSerSerArgSer 212
         |||||||||||||||||||||||||||||||||||||||
Db   616 TGTCGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 654
```

RESULT 6
ABX92075
ID   ABX92075 standard; cDNA; 1977 BP.
XX
AC   ABX92075;
XX
DT   08-MAY-2003 (first entry)
XX
DE   Lung specific nucleic acid (LSNA) #117.
XX
KW   Human; gene therapy; vaccine; lung specific antigen; cancer diagnosis;
KW   cancer monitoring; cancer staging; cancer imaging; lung cancer;
KW   non-cancerous diseases of the lung; transgenic animal; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20026833-A2.
XX
PD   06-SEP-2002.
XX
PF   21-NOV-2001; 2001WO-US043612.
XX
PR   22-NOV-2000; 2000US-0252500P.
XX

PA   (DIAD-) DIADEXUS INC.
XX
PI   Macina RA, Recipon H, Chen S, Sun Y, Liu C;
XX
DR   WPI; 2002-713376/77.
XX
PT   New isolated human nucleic acid molecule and polypeptide, useful for
PT   identifying, diagnosing, monitoring, staging, imaging and treating lung
PT   cancer and non-cancerous diseases of the lung.
XX
PS   Claim 1; Page 266-267; 389pp; English.
XX
CC   The invention describes an isolated human nucleic acid (I) encoding any
CC   of 120 10-1533 residue amino acid sequences (S1), given in the
CC   specification, comprising any of 164 179-12421 base pair sequences (S2),
CC   given in the specification. The methods and compositions of the present
CC   invention are useful for identifying, diagnosing, monitoring, staging,
CC   imaging and treating lung cancer and non-cancerous diseases of the lung.
CC   They are also used for identifying lung tissue, monitoring and
CC   identifying and/or designing antagonists of the polypeptide of the
CC   invention, gene therapy, production of transgenic animals and production
CC   of engineered lung tissue for treatment and research. This sequence
CC   encodes a lung specific nucleic acid
XX
SQ   Sequence 1977 BP; 393 A; 631 C; 554 G; 399 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   1.18e-30   Length: 1977
Score: 1002.50   Matches: 206
Percent Similarity: 94.5%   Conservative: 2
Best Local Similarity: 93.6%   Mismatches: 4
Query Match: 88.4%   Indels: 8
DB: Gaps: 1

US-09-989-890-238 (1-212) x ABX92075 (1-1977)

Qy   1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 AGCCCCCACCAAGCCCGCCACCCGTAGACCCAAGACCCCAAGACCCTGGCCACCATGGGC 416

Qy   21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 CAGAGAGCATTACCTTACCCTCTGGCCTCGCTGCTGAGCCCCTTGAGTCCCCCACCTGCT 476

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla-Al 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| :::
Db   477 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCCCGC 536

Qy   60 aAlaGlyIleAlaSerSerAlaValGlu-ProValCysGlyAspAlaAlaProAlaCysL 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 TGCCGGGATTGCCTCCAGCGCTCGTGTGCCCGTGGGGGATGCAGCCCCTGCCTGTC 596

Qy   80 euLeuArgThr-ProLeuArgGlyLeuLeuLysProThrGly-ProArgSerThrMetGl 99
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   597 TACTGAGGACTCCCACTGAGGGACTGCTGAAGCCAACTGGTGCCAAGGAGCACAATGA 656

Qy   99 uCysProProAlaLeuIleValHisProPro-AlaGlyGlyMetAla---SerGlySerS 118
          |||||||||||||||||||||||||||||||||||||||||||| :::
Db   657 GTGCCCCCCAGCCCTGATCGTGCACCCCCAGACGCGGGCGGATGGCCAGGCGGGCTGCAA 716

Qy   118 erGlnProTrpAlaAlaThrProAlaAlaThrProMetLeuSerSerLysAlaSerLeuCysI 138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   717 GTCAACATGGGGACGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTA 776

Qy   138 leProThrArg-GlyProProProGlnProLeuMetArgThrProAlaAlaArgSerHis 157
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   777 TCCCTACCCGGAGAGGACGCCACCTCCCCAGCCCTGATGCGGGACTCCTGCTGCAAGGAGCCAC 836

Qy   158 TrpProIleProHisProCysAsp-ThrAlaCysProLeuProLeuProValValLeuVa 177
          ||||||||||||||:::|||||||||||||||||||||||||||||
Db   837 TGGCCGATCCCCCACCCATGCGAGCACAGACCTGCCCAGCACCTTTGCCAGTAGTCCTCGT 896

Qy   177 lAlaProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaVa 197
          |||||||||||||||||||||||||||||||||||||||||

```
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences.
XX
PS   Claim 9; Page 854; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotide. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia,
XX
SQ   Sequence 427 BP; 69 A; 161 C; 123 G; 73 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:             1.5e-08         Length:           427
Score:                 417.00          Matches:          81
Percent Similarity:    100.0%          Conservative:     0
Best Local Similarity: 100.0%          Mismatches:       0
Query Match:           36.8%           Indels:           0
DB:                    5               Gaps:             0

US-09-989-890-238 (1-212) x AAF66376 (1-427)

Qy   1   SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183 AGCCCCCACCAAGCCGGCGCACCCGTAGACCAAGACCCCAAGGACCCTGGCCACCATGGGC 242

Qy   21  GlnArgAlaAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProPro Ala  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 CAGAGAGCATTACCTTCATCTCGGCTCTGCTGAGCCGGCGCCTTGAGTCCCCACCTGCT   302

Qy   41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303 GCCTGCTCTGGCGACCCTGGGTGTGGAGGTGGTGCCGGGCTGCCTTCTCCGCCGCT     362

Qy   61  AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 GCCGGGATTGCCTCCAGCGCGTGTGGAGCCTGTTGCGGNGATGCAGCCCCTGCCTGTCTA  422

Qy   81  Leu  81
         |||
Db   423 CTG  425
```

RESULT 9
ACL72571/c

```
ID   ACL72571 standard; DNA; 1716 BP.
XX
AC   ACL72571;
XX
DT   02-JUN-2005 (first entry)
XX
DE   M. xanthus gene sequence, seq id 9034.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression;
KW   gene; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
```

```
Db   417 CAGAGAGCATTACCTTCATCTCGGCTCTGCTGAGCCGGCGCCCTTGAGTCCCCACCTGCT 476
Qy   41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla-Al 60
Db   477 GCCTGCTCTGGCGACCCTGGGTGTGGAGGTGGTGCCGGGCTGCCTTCTCCGCTTCCGCCCGC 536

Qy   60  aAlaGlyIleAlaSerSerAlaValGlu-ProValCysGlyAspAlaAlaProAlaCysL 80
Db   537 TGCCGGGATTGCCTCCAGCGCGTGTGGAGGCCTGTTGCGGGGATGCAGCCCCTGCCTGTC 596

Qy   80  euLeuArgThr-ProLeuArgGlyLeuLeuLysProThrGly-ProArgSerThrMetGl 99
Db   597 TACTGAGGACTCCCACTGAGGGGACTGCTGAAGCCAACTGGTGCCAAGAGCACAATGGA 656

Qy   99  uCysProProAlaLeuHisProPro-AlaGlyGlyMetAla---SerGlySerS 118
Db   657 GTGCCCCCAGCCCCTGCATCGTGCACCCCCAGACCGGCGGATGGCCAGCGGGCTGCAA 716

Qy   118 erGlnProTrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysI 138
Db   717 GTCAACCATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTA 776

Qy   138 leProThrArg-GlyProProProGlnProLeuMetArgThrProAlaAlaArgSerHis 157
Db   777 TCCCTACCCGGAGAGGCCACCTCCCCAGCCCCTGATGCGACTCCTGCTGCAAGAGAGCCAC 836

Qy   158 TrpProIleProHisProCysSerAsp-ThrAlaCysProAlaProLeuProValValLeuVa 177
Db   837 TGGGCCGATCCCCCACCCATGCGAGCACAGCCTGCCCCAGCACCTTTGCCCAGTAGTCCTCGT 896

Qy   177 lAlaProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaVa 197
Db   897 GGCTCCGAGGAGTACTATTCTTTCCATGAGTCGGACCTGCCGGGAGATGGGCAGT 956

Qy   197 lAlaProCysArgAlaGluGlyLysLeuMetCysSerSerSerArgSer 212
Db   957 GGCTCCATGTCGAGCGCGGAGAAATTGATGTGCTCATCTTCAAGAAGC 1002
```

RESULT 8
AAF66376

```
ID   AAF66376 standard; cDNA; 427 BP.
XX
AC   AAF66376;
XX
DT   09-APR-2001 (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 2132.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US018374.
XX
PR   02-JUL-1999; 99US-0142310P.
PR   02-JUL-1999; 99US-0142311P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J,
PI   Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;
PI   Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;
PI   Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
```

```
Db   1027 --CGTCCGAGGAGGAAGGCCATCAGCAACGAGGGGGTCCCGACGCGGCAGCGGCGGCGGCCCG  971
          ||||                                                    ::::
Qy    141 ----------------------ArgGlyProProGlnPro-------------------- 147
                                ||||||||:::||
Db    970 TGTTCAGCGTGGGCGGCGAGGCAACGTAGAGGTCCCCACCGGGTCCGAGGCGGAGGCCAT  911
                                                                  |||||
Qy    148 ------------------------------------------LeuMetArgThrPro    152
                                                    ||||||
Db    910 ACGGGTGCGGTGGTTGCACTCGGCGACATCGGTCCGTCAACTCGCGCGGACACCA       851
                                                         |||||||
Qy    153 AlaAlaAlaArgSerHisTrpProIlePro-----HisProCysAspThrAlaCys----- 168
          |||     |||      |||||                    |||
Db    850 TGTGGCCGGAGGCAT---CCAGCACCACCAGGAGTACATGCGGCACGCCGTGTCGCGCC   794
                          |||||||||                         |||||
Qy    169 ----------------------------------------ProAlaProLeuProVal    174
                                                    |||||||||
Db    793 GCTGGTCCGTGGGTAGGCTCGAAGAGGCCGCCAGCCGGTCCGGCGCCATCACCGCCA    734
                                                 |||||       |||
Qy    175 ValLeuValAlaProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArg  194
                                       |||||       :::
Db    733 ------GCCGGCGCAGGAGAGCGTCTGCGTCCTGTCCCGGAAGTCGGGCGCGCGCGTCGG  680
                ||||||                       :::
Qy    195 --------------------TrpAlaValAlaProCysAArgAla               202
                              |||||||||||
Db    679 CCAGGGTGGAGGTCCACCGCGGTGTCCCCGTGTCGCGCT               641
```

```
RESULT 10
ACL64503/c
ID   ACL64503 standard; DNA; 9198 BP.
XX
AC   ACL64503;
XX
DT   02-JUN-2005  (first entry)
XX
DE   M. xanthus DNA fragment, seq id 966.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,   Hinkle GJ,   Slater SC,   Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 1; SEQ ID NO 966; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC   set of 1849 contig and singleton sequences comprising coding sequences,
CC   DNA replication elements, promoters and other regulatory elements from
CC   the genome of the bacterium Mxyococcus xanthus. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
```

```
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS,   Hinkle GJ,   Slater SC,   Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 2; SEQ ID NO 9034; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC   set of about 7842 genes or partial genes from the genome of the bacterium
CC   Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from USPTO
XX
SQ   Sequence 1716 BP; 230 A; 580 C; 660 G; 246 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:             85                    Length:       1716
Score:                 178.00                Matches:      84
Percent Similarity:    33.1%                 Conservative: 13
Best Local Similarity: 28.7%                 Mismatches:   88
Query Match:           15.7%                 Indels:       108
DB:                    14                    Gaps:         15
```

```
US-09-989-890-238 (1-212) x ACL72571 (1-1716)

Qy      2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln  21
          |||||||                 |||
Db   1471 CCCCATCCGGCGCGGCA--------------------CGCCCACCAGCAACCGCTGGTTGT 1430

Qy     22 ArgAlaLeuProSerLeuLeuLeuLeuSerArgProLeuSerProProProAlaAla  41
          |||                    ::|                     |||||||
Db   1429 CGGCGCGCCACCGCGTGCGTCAACACCAGCGTCTCGGCGGCGGAAGCCCCGCTGGGCC 1370

Qy     42 -----CysSerGlyAspProGlyCysGlyGlyValGlyGlyLeuProSerAlaSerAla  59
               :::||                  |||||||||||||||
Db   1369 AGGTGGCCAGCCGGAGCTCCT--TGCTCGTCAGGTGCTGGCCTTCCTTCAGCGAGTACA 1313

Qy     60 Ala-----------------AlaGlyIleAlaSerSerAlaValGluProValCys  72
          |||                 |||||      ||
Db   1312 GCCACGCGCGGCTCCCAGGTCCCGGAGCGGAAGCCCTTCAGCTGCGTGGCCGGCGTGA 1253

Qy     73 GlyAspAla-----------AlaProAlaCysLeuLeuArg---ThrProLeuArg  86
          |||                 |||      :::              |||
Db   1252 AACTCAGGCTGCGTCCACATCCGCGTTCACCGTGGTGCCCACGCGCTGGGCACCACCGT 1193

Qy     87 GlyLeuLeuLysProThrGlyProArgSerThrMetGluCysProProAlaLeuIle-- 105
          |||                     |||                    :::
Db   1192 CCCGGGTGGTGGCCACCGCGCGGCCCAGTTGCTGGGCGGAGGTGCTCCCCACCGGTGTTCCGT 1133

Qy    105 -------------------------------------------------------- 105
Db   1132 CCCGGGTGCTCAACGCCGCTGGAAGCCCATTCGGCAGCAGCAGCCCCGTTCACCACCGCGA 1073

Qy    106 ValHisProAlaGlyGlyMetAlaSerGlyGlySerGlnProTrpAlaAlaAlaSer 125
          |||:|||                                 :::
Db   1072 GCTCGCCGCGCGCAGGGGCGTCCCCGCCAGCGCGGCGCACCC------------ 1028

Qy    126 AlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr-------- 140
```

CC   obtained in electronic format directly from USPTO
XX
SQ   Sequence 9198 BP; 1247 A; 3175 C; 3389 G; 1387 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   344   Length: 9198
Score: 178.00   Matches: 84
Percent Similarity: 33.1%   Conservative: 13
Best Local Similarity: 28.7%   Mismatches: 88
Query Match: 15.7%   Indels: 108
DB: 14   Gaps: 15

US-09-989-890-238 (1-212) x ACL64503 (1-9198)

Qy   2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
          |||||| ||| ||| |||||| |||||| |
Db 1471 CCCCATCCCGCGGCGCA----------------------------CGCCCACCGACCGCTGGTTGT 1430

Qy   22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAlaAla 41
          |||||| ::: ||||||||| |
Db 1429 CGGCGGCCGCCACCGCGTCGCTCAACACCAGCCGTCTCGGCGGAAGCCCCGCCCCCTGGGCC 1370

Qy   42 ------CysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla 59
                :::||| ||| |||
Db 1369 AGGTGGCCAGCCGAGCCTCCT--TGCTCGTCAGGTGCTGGCCTTCCTTCAGCCAGTACA 1313

Qy   60 Ala------------AlaGlyIleAlaSerSerAlaValGluProValCys 72
          ||| ||| :::||| |||
Db 1312 GCCACGCCGGCTCCAGGTCCGGGAGCCGGAAGCCCTTCAGCTGGGTGGCGGCGGCGTGA 1253

Qy   73 GlyAspAla------------AlaProAlaCysLeuLeuArg---ThrProLeuArg 86
          :::||| |||||||| |||
Db 1252 AACTCAGGCTGTCCACATCCGCGTTCACCGTGGTGCCCGACGCGGCTGGGCACCACGT 1193

Qy   87 GlyLeuLeuLysProThrGlyProArgSerThrMetGluCysProProAlaLeuIle--- 105
          |||||| ||| || :::
Db 1192 CCCGGGTGGCCACCGCGCCGCCCAGTTGCTGGGCGAGGTGCTCCCCCACCGGTGTTCCGT 1133

Qy   105 105
Db 1132 CCCGGGTGCTCAACGCCTGAAGCCCCATTCGGCGGCAGCAGCAGCAGCCCGTTCACCACCGGGA 1073

Qy   106 ValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnProTrpAlaAlaAlaSer 125
          :::||||||||||| :::|||
Db 1072 GCTCGCCCCGCGCGAGGGGGCTCCGTCCCTCCGCAGCCGCCACCC--------------- 1028

Qy   126 AlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr--------------- 140
          ||||| ||| ||| :::
Db 1027 ---CGTCCGAGGAGAAGCCCATCAGCAACGACGAGGGCGTCCCGACGCCAGCGCGGCGCCCG 971

Qy   141 ------ArgGlyProProGlnPro--- 147
                |||||||||||::|||
Db   970 TGTTCAGCGTGGGCGCGGAAGGCAACGTAGAGGTCCCCACCGCGTCCGAGGCGAGGCCAT 911

Qy   148 ---------------------------LeuMetArgThrPro 152
                                     |||||||
Db   910 ACGGGTGCGGCGTGGTTGCACTCGGCCGCGAGCATCGGGTCCGTCAACTCGCGCGGCGACACCA 851

Qy   153 AlaAlaArgSerHisTrpProIlePro------HisProCysAspThrAlaCys----- 168
          |||||| ||| ||| |||
Db   850 TGTGGCCGAAGGCAT---CCAGCACCACCAGGAGTACATGCGGCACGCCGTGTTCGCGCC 794

Qy   169 ------ProAlaProLeuProVal 174
                ||||||||||||
Db   793 GCTCGTCGTCGTGGGGTAGGCCTCGAGAGGGCGCCCAGCCGGCCGGCCGCCATCACCCGCCA 734

Qy   175 ValLeuUalValAlaProSerThrIleLeuSerMetSerArgThrTrpThrCysArgArg 194
          ||| ::: |||
Db   733 ------GCCGGCCAGGAAGAGAGCCTCTGGCTCCTGTCCCGGAAGTCCCGGCCGCCGGTCGG 680

Qy   195 ---------------TrpAlaValAlaProCysArgAla 202
                         |||||| :::|||||||
Db   679 CCAGGGGTGAAGGTCCACCGCGGTGTCCCCGTGTCGCGCT 641

```
DT   31-JUL-1991  (first entry)
XX
DE   Eimeria tenella 5401 sporozoite antigen gene.
XX
KW   Antigen; vaccine; poultry; fowl; avian coccidiosis; immunization; ss.
XX
OS   Eimeria tenella.
XX
PN   WO8600528-A.
XX
PD   30-JAN-1986.
XX
PF   05-JUL-1985;   85WO-US001274.
XX
PR   05-JUL-1984;   84US-00627811.
PR   19-JUN-1985;   85US-00746520.
XX
PA   (GENEX ) GENEX CORP.
XX
PI   Anderson DM, Mccandliss RJ;
XX
DR   WPI; 1986-042037/06.
DR   P-PSDB; AAP60577.
XX
PT   New cloned gene expressing Eimeria antigen - useful in vaccines for
PT   protecting poultry against coccidiosis.
XX
PS   Claim 7; Page 68-70; 82pp; English.
XX
CC   The sequence codes for an antigenic protein which binds with an antibody
CC   directed against an antigenic protein of avian coccidia, (esp. a protein
CC   of B.tenella sporozoite). It may be used for immunising birds against
CC   avian coccidiosis. The specification also includes a generic version of
CC   this sequence to cover all the degenerate codons coding for a protein
CC   having the amino acid sequence of AAP60577. (Updated on 25-MAR-2003 to
CC   correct PA field.)
XX
SQ   Sequence 864 BP; 187 A; 180 C; 385 G; 112 T; 0 U; 0 Other;
```

Alignment Scores:
```
Pred. No.:              107
Score:                  169.00
Percent Similarity:     38.7%        Length:        864
Best Local Similarity:  28.9%        Matches:        56
Query Match:            14.9%        Conservative:   19
DB:                     1            Mismatches:     91
                                     Indels:         28
                                     Gaps:           5
```

US-09-989-890-238 (1-212) x AAN60488 (1-864)

```
Qy    2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
          |||          |||||||          ::::
Db  710 CCGGAGTACGCGGCGGCAACGGCCA----------CCACCCACTGCTGCCAGCAGTAGTACA 663

Qy   22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAlaAla 41
            |||         ::: |||||:::
Db  662 CCACCTACACCTCCCGGCAACAGCTGCCGTGGGGAAGCCGCTGCCTCCTTCCTCTCT 603

Qy   42 CysSerGlyAspProGlyCysGlySerGlyArgAlaGlyLeu----------ProSer 56
          |||||          |||||
Db  602 TCTTCTGCACCGCCTTCGGTGGCTCCTCCTGCTGGACGGGTTCCGTCCGAGGGTTGCTTCA 543

Qy   57 AlaSerAlaAlaAlaGlyIleIleAlaSerSerAlaValGluProValCysGlyAspAlaAla 76
          :::|||||::::::          |||
Db  542 GGTCTCTCAGCTTCAGGTGTTTCCGGTTCGGGTGTCTCAGGGGTTTCGGCCACTTGCTCG 483

Qy   77 ProAlaCysLeuLeuArgThrProLeuLeuLeuSerArgGlyLeuLeuLysProArgGlyProArgSer 96
          |||          |||:::|||
Db  482 CCTCCTCCACTGCAGCCCCCTTCCCTCCTCCCAGGCTGGGACCTGACGTCGAGACCTTCACCGGCCAGCA 423

Qy   97 ThrMetGluCysProProAlaLeuIleValHisPro---------- 108
          ||||||
Db  422 CCCTCACCCAGCGCCGCCTGCCGCCACCCCTTCCGCCCGCAGTTTCCCGTCGTCGCCTTCG 363
```

Alignment Scores:
```
Pred. No.:              206
Score:                  173.00
Percent Similarity:     34.3%        Length:        2914
Best Local Similarity:  25.8%        Matches:        70
Query Match:            15.3%        Conservative:   23
DB:                                  Mismatches:     88
                                     Indels:         90
                                     Gaps:           12
```

US-09-989-890-238 (1-212) x ADJ92146 (1-2914)

```
Qy    1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
          :::|||    |||||||
Db 2194 GCCCCTGCCAGTCAGGCTGCCACCAGCCTGCCCCTGCT----------- 2238

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40
                        |||||| |||
Db 2239 -------GCCAGAGTCTAGCTGCCAGCCAGCTTGCTGCACATCCTCCCCCTGCC 2286

Qy   41 ----------AlaCysSerGlyAspProGlyCysGlySerGlyAla 52
                    |||   |||||||
Db 2287 AGCAGGCCTGCTGCGTGCCTGTCGCAAGCCAGTCTGTGTGCCCGTCTGCTGCA 2346

Qy   53 GlyLeuProSerAlaSer 58
          |||  :|||||
Db 2347 AGCCTGTCTGCGCAAGCCCATCTGGTCGCGTCTGCGCTGCTGGGGCTTCCTCTTCAT 2406

Qy   59 AlaAlaAlaGlyIleAlaSerSer------AlaValGluProValCysGlyAspAlaAla 76
          |||:::|| ||||:::||||
Db 2407 GCTGCCAGCAGTCAGGTTGCCCAGCCGGCTTGCTGCACCACCTCCTGCTGCAGACCCCTCT 2466

Qy   77 ProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLysProThr----------- 92
          ||| |||
Db 2467 CCTCCGTGT-----------------------------CCCTCCTCTGCCGCCGCCCCG 2493

Qy   93 -----GlyProArgSerThrMetGluCysProProAlaLeuIleValHisProProAla 110
               |||||||||    ::: ||||||||||||
Db 2494 TGTGCAGTGCAGTGCCACCTGCTGTGTGCCCATCCCCTGTGCCCCTGCCCACCTGCC 2553

Qy  111 GlyGlyMetAlaSerGlySerGlnProTrpAlaAlaAlaSerAlaThrProMetLeu 130
          ::: ||| |||||
Db 2554 --AGCCCAGCTGCTGCCGCCCCGACCCCTGCCCCTGCTGTCCCCTCCGCCCCA----- 2604

Qy  131 SerSerLysAlaSerLeuCysIleProThrArgGlyProProGlnProLeuMetArg 150
          |||:|| :::|||
Db 2605 ------CGTGCTCCCGGCCTCTTCCCGCGGCTGCGGCCTCTTCCTCAG------GCCAGA 2652

Qy  151 ThrProAlaAlaArgSerHisTrpProIleProHisProCysAspThrAla------- 167
          :::|||||||||
Db 2653 AGTCCAGCTGCTGACGGGCGAGGAGTTCCCGTTTCTCCAGAGTGTGTTCCAGGGCCATCCTCTGAC 2712

Qy  168 ----------CysProAlaProLeuProValValLeuVal 177
                     ::: ||||:::
Db 2713 TTCCGAAAATCCTTGCAGCCCTGAGCCCTGTCTGAGTCCCATAGCCACTGCA 2772

Qy  178 AlaProArgSer------------ 181
          :::| |||
Db 2773 CCTCCAGCCAGCAATGCTGAGCTTTCATGCCAATCCCGATTCCCAGGCAGCATGCATCTCTCCCAC 2832

Qy  182 ThrIleLeuSerMetSerArgThrThrCysArgArgTrpArgAlaValAlaProCysArg 201
          |||:::::|||||| |||
Db 2833 ACCCTCATCTCCATG---------------CTGGCCAGCCGGTGGTCACTGTGTCCTGCCCCA 2880

Qy  202 AlaGluLysMetCysSerSerSerArgSer 212
          :::||||||
Db 2881 TGTGAGTGATCAGTCGTGTCAACAATAAGTCA 2913
```

```
RESULT 12
AAN60488/c
ID   AAN60488 standard; DNA; 864 BP.
XX
AC   AAN60488;
XX
DT   25-MAR-2003 (revised)
```

```
Score:                  169.00     Matches:         56
Percent Similarity:      38.7%     Conservative:    19
Best Local Similarity:   28.9%     Mismatches:      91
Query Match:             14.9%     Indels:          28
                                   Gaps:             5
DB:

US-09-989-890-238 (1-212) x AAQ03317 (1-864)

Qy    2  ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln    21
         ||||||        ||||||          |||  |||
Db  710  CCGGAGTACGGCGGCCAACGCA----------CCACCCACTGCTGCCAGCAGTAGTACA   663

Qy   22  ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAlaAla    41
                           |||        |||    :::::::
Db  662  CCACCTACACCTCCGCAACAGCTGCGCTGGGGAAGCCGCGTGCCCCTCCTCCTCCTCCT   603

Qy   42  CysSerGlyAspProGlyCysGlySerGlyAlaGlyLeu------------ProSer     56
         |||||         |||||||           ||::||:
Db  602  TCTTCTGCACCGCCTTCGGTGGGCTCCTCTGCTGGAGTTTCCGTCGAGGGTTGCTCTTCA   543

Qy   57  AlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAla    76
            :::::||::||::|||:::::::      |||       |||
Db  542  GGTCTCCTCAGCTTCAGGTGTTCCGGTTCGGGTGTCTCAGGGGTTCCGGCACTTGCTCG   483

Qy   77  ProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLysProGlyThrGlyProArgSer  96
                     :::|||         ||:||
Db  482  CCTCCTCACTCGCCCCTTCTCCCTCCCCAGGCTGGACCTGACCTTCACCACGCCAGCA   423

Qy   97  ThrMetGluCysProProAlaLeuIleValHisPro------------------------  108
                                   |||||
Db  422  CCCTCACCAGGCCGCCGGCTGCGCCACCCCTTCGCCGCCGCCAGGTTCCCGTCGTCCGCCTTCG   363

Qy  109  ------------ProAlaGlyGlyMetAlaSerGlySerSerGln------------------  119
                       |||||||                    ||||||||
Db  362  GCTCCTCCAGCACCTGCGCGCCGCCCCTCTTCCCGGGCAGCTTCCTCAGCATTGCCGCCAGAG   303

Qy  120  ---ProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIle    138
                |||                                   :::|||
Db  302  CCTCCTTCTCCGCGCCGCACCTGCGCCGCCCCTCTTCTCCAGGCAGCTCCTCAGCACTACCG   243

Qy  139  ProThrArgGlyProProProGlnProLeu--MetArgThrProAlaAlaArgSerHis    157
         |||||                            :::||||          :::|||||
Db  242  CCAGAGCGTCCTCCTCCGCCGGCACCTGCGCCGCCCCTCTTCTCCGGGCAGCTCCTCAGCA   183

Qy  158  TrpProIleProHisProCysAspThrAlaCysProAlaPro    171
         |||||                          ||:::|||
Db  182  CTACCGCCAGAGCCTCCTCCAGCTTCGCCAGCACCCCCTTCTCCG   141


RESULT 14
AAQ03324/c
ID   AAQ03324 standard; DNA; 6567 BP.
XX
AC   AAQ03324;
XX
DT   25-MAR-2003  (revised)
DT   02-AUG-1990  (first entry)
XX
DE   Eimeria tenella genomic DNA encoding antigen GX5401FL.
XX
KW   Eimeria tenella; antigen GX5401FL; antigen GX5401; avian coccidiosis.
XX
OS   Eimeria tenella.
XX
FH   Key            Location/Qualifiers
FH
FT   CDS            1..6567
FT                  /*tag= a
XX
PN   WO900403-A.
XX
PD   25-JAN-1990.
XX
PF   05-JUL-1989;  89WO-US002918.
XX
PR   05-JUL-1989;  89WO-US002918.
XX
```

```
Qy  109  ------------ProAlaGlyGlyMetAlaSerGlySerSerGln------------------  119
                       |||||||                    ||||||||
Db  362  GCTCCTCCAGCACCTGCGCGCCGCCCCTCTTCCCGGGCAGCTTCCTCAGCATTGCCGCCAGAG   303

Qy  120  ---ProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIle    138
                |||                                   :::|||
Db  302  CCTCCTTCTCCGCGCCGCACCTGCGCCGCCCCTCTTCTCCAGGCAGCTCCTCAGCACTACCG   243

Qy  139  ProThrArgGlyProProProGlnProLeu--MetArgThrProAlaAlaArgSerHis    157
         |||||                            :::||||          :::|||||
Db  242  CCAGAGCGTCCTCCTCCGCCGGCACCTGCGCCGCCCCTCTTCTCCGGGCGGCTCCTCAGCA   183

Qy  158  TrpProIleProHisProCysAspThrAlaCysProAlaPro    171
         |||||                          ||:::|||
Db  182  CTACCGCCAGAGCCTCCTCCAGCTTCGCCAGCACCCCCTTCTCCG   141


RESULT 13
AAQ03317/c
ID   AAQ03317 standard; DNA; 864 BP.
XX
AC   AAQ03317;
XX
DT   25-MAR-2003  (revised)
DT   02-AUG-1990  (first entry)
XX
DE   cDNA of Eimeria tenella oocyte encoding antigens GX5401 and GX3264.
XX
KW   Eimeria tenella; antigen GX5401; antigen GX3264; avian coccidiosis.
XX
OS   Eimeria tenella.
XX
FH   Key            Location/Qualifiers
FH
FT   CDS            1..864
FT                  /*tag= a
FT   misc_feature   70..864
FT                  /*tag= b
FT                  /note= "DNA sequence of GX3264"
XX
PN   WO900403-A.
XX
PD   25-JAN-1990.
XX
PF   05-JUL-1989;  89WO-US002918.
XX
PR   05-JUL-1988;  88US-00215162.
XX
PA   (GEMX ) GENEX CORP.
XX
PI   Anderson D, Mccandliss R;
XX
DR   WPI; 1990-051586/07.
DR   P-PSDB; AAR05216.
XX
PT   Cloned gene or fragment encoding antigenic protein - which binds with
PT   antibodies against avian coccidia, and transformed cells used in vaccine.
XX
PS   Claim 1; Page 93 and Fig 1; 134pp; English.
XX
CC   It encodes two antigenic proteins (AP) which bind with a monoclonal
CC   antibody (MAb) or polyvalent Ab directed against an AP of avian coccidia.
CC   GX5401 is claimed in claim 1, and GX3264 is claimed in claim 2. Also new
CC   are an expression vector contg. the cloned gene, host cells transformed
CC   with the vector, and AP encoded by the cloned gene. The transformed cells
CC   are used in a vaccine to immunise birds against avian coccidiosis. By
CC   labelling the peptides, they can be used as a type-specific probe. The AP
CC   may also be used in an assay to detect Ab against the coccidia. The Abs
CC   are used to identify transformed cells contg. the DNA. (Updated on 25-MAR
CC   -2003 to correct PA field.)
XX
SQ   Sequence 864 BP; 187 A; 180 C; 385 G; 112 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:       107            Length:        864
```

```
Qy    120  ---ProTrpAlaAlaAlaSerAlaAlaThrProMetLeuSerSerLysAlaSerLeuCysIle 138
                  |||      |||   |||  |||  ||:::|||
Db   6008  CCTCCTTCCGCCGGCACCTGCGCCGCCGCCTCCTTCCCAGGCAGCCAGCTCCTCAGCACTACCG 5949

Qy    139  ProThrArgGlyProProGlnProLeu---MetArgThrProAlaAlaArgSerHis 157
                  |||||||||||   :::  |||        |||  ::: |||
Db   5948  CCAGAGCCTCCTGCCCGCACCTGCGCGCGCCCTTCTCCCGGCAGCTCCTCAGCA 5889

Qy    158  TrpProIleProHisProCysAspThrAlaCysProAlaPro 171
                  |||                     |||:::||
Db   5888  CTACCGCCAGAGCCTCCAGCTTCGCCCAGCACCCCCTTCTCCG 5847
```

RESULT 15
AB22978/c
```
ID   AB22978 standard; DNA; 6567 BP.
XX
AC   AB22978;
XX
DT   10-APR-2003  (first entry)
XX
DE   Eimeria maxima 250 kDa protein homologous DNA sequence SEQ ID NO:5.
XX
KW   Eimeria maxima; 250 kDa antigen; sporozoite; merozoite; infection;
KW   vaccine; antiparasitic; gene therapy; immunodominant; gene; ds.
XX
OS   Eimeria maxima.
OS   Synthetic.
XX
PN   WO200304684-A2.
XX
PD   16-JAN-2003.
XX
PF   03-JUL-2002; 2002WO-US021237.
XX
PR   06-JUL-2001; 2001US-0303670P.
XX
PA   (WITC/) WITCOMBE D.
PA   (SMIT/) SMITH N C.
PA   (WALL/) WALLACH M.
XX
PI   Witcombe D, Smith NC, Wallach M;
XX
DR   WPI; 2003-201556/19.
XX
PT   New nucleic acid comprising a sequence encoding a 250 kDa polypeptide
PT   from sporozoites/merozoites of Eimeria maxima, useful for preparing a
PT   vaccine against Eimeria infection.
XX
PS   Example 3; Fig 7A; 198pp; English.
XX
CC   The present invention describes a nucleic acid (I) comprising a sequence
CC   encoding a 250 kDa polypeptide from Sporozoites/Merozoites of Eimeria
CC   maxima or its homologue or complement. Also described: (1) a vector
CC   comprising the nucleic acid; (2) a host cell comprising the vector; (3) a
CC   plasmid comprising the nucleic acid; (4) a transformed cell comprising
CC   the nucleic acid; (5) producing a recombinant 250 kDa polypeptide; (6) a
CC   recombinant polypeptide; (7) a vaccine against E. tenella, E. acervulina,
CC   E. necatrix, E. praecox, E. maxima, E. mitis or E. brunetti or a
CC   microorganism expressing an immunologically cross-reactive antigen; (8)
CC   immunising a subject against infection by Eimeria, or a microorganism
CC   expressing an immunologically cross-reactive antigen; (9) conferring upon
CC   a newborn subject of an avian species maternal immunity against infection
CC   by Eimeria; (10) a fertilised egg from an avian species having an air
CC   sac, where the air sac is inoculated with the vaccine; and (11) reducing
CC   the output of Eimeria oocytes in faeces from a newborn subject of an
CC   avian species. (I) has antiparasitic activity and can be used in a
CC   vaccine and in gene therapy. The nucleic acid is useful for preparing a
CC   vaccine against E. tenella, E. acervulina, E. necatrix, E. praecox, E.
CC   maxima, E. mitis or E. brunetti infection. The present sequence
CC   represents a 250 kDa antigen protein homologous DNA sequence, which is
CC   used in an example from the present invention
XX
```

```
PR   05-JUL-1988;  88US-00215162.
XX
PA   (GEMX ) GENEX CORP.
XX
PI   Anderson D, Mccandliss R;
XX
DR   WPI; 1990-051586/07.
DR   P-PSDB; AAR05222.
XX
PT   Cloned gene or fragment encoding antigenic protein - which binds with
PT   antibodies against avian coccidia, and transformed cells used in vaccine.
XX
PS   Claim 10; Page 93 and Fig 14; 134pp; English.
XX
CC   An E. tenella genomic library was screened with radioactively labeled
CC   cDNA encoding the GX5401 antigen to identify clones encoding extensions
CC   of the GX5401 coding sequence. Several plaques reactive with the cDNA
CC   were identified. DNA from clone number 533 was sequenced. This analysis
CC   showed that the coding sequence in this clone overlaps with the GX5401
CC   coding sequence and extends toward the N-terminal coding
CC   region. Analysis of the new sequence together with the GX5401 antigen
CC   coding sequence reveals an open reading frame encoding an Eimeria protein
CC   of about 250,000 daltons. The protein carries several repeated peptide
CC   sequences and is rich in cysteine residues. The open reading frame
CC   encodes a potential signal sequence for protein secretion. The coding
CC   sequence for this antigenic protein is given in AAN9324. Also new are an
CC   expression vector contg. the cloned gene, host cells transformed with the
CC   vector, and AP encoded by the cloned gene. The transformed cells are used
CC   in a vaccine to immunise birds against avian coccidiosis. By labelling
CC   the peptides, they can be used as a type-specific probe. The AP may also
CC   be used in an assay to detect Ab against the coccidia. The Abs are used
CC   to identify transformed cells contg. the DNA. (Updated on 25-MAR-2003 to
CC   correct PA field.)
XX
SQ   Sequence 6567 BP; 1490 A; 1741 C; 2189 G; 1147 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:                              Length:        6567
Score:                    580           Matches:       56
Percent Similarity:       169.00        Conservative:  19
Best Local Similarity:    38.7%         Mismatches:    91
Query Match:              28.9%         Indels:        28
DB:                       14.9%         Gaps:          5
                          2
```

```
US-09-989-890-238 (1-212) x AAQ03324 (1-6567)

Qy      2  ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
                  |||       |||                 |||  |||::
Db   6416  CCGGAGTACGCGGCAACGCCA------------CCACCCACTGCTGCCAGCAGTAGTACA 6369

Qy     22  ArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAlaAla 41
                  :::     |||   ::: |||              |||:::||||||
Db   6368  CCACTACACCTCCGGCAACAGCTGCCGTGGGAGGAAGCCGGCCGCTCCTTCCTCCTCT 6309

Qy     42  CysSerGlyAspProGlyCysGlyCysGlyAlaGlyLeu----------ProSer 56
                  ||||||   |||  ::: |||                      |||
Db   6308  TCTTCTGCACCGGCCTTCGGTGGGCTCCTCGCTGGAGTTTCCGTCGAGGGGTTGCTCTTCA 6249

Qy     57  AlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAla 76
                  :::||||||||::::|||   |||  ||    |||        :::
Db   6248  GGTCTCTCAGCTGTTTCCGGTTCGGTGTCTCAGGGGTTTCCGGCACTTGCTCG 6189

Qy     77  ProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSer 96
                  |||                 |||  |||  |||     :::  :::
Db   6188  CCTCCTTCACTGCCCCCTTCCTCCTCCAGGCTGGACCTCACCACCTTCACCGCCAGCA 6129

Qy     97  ThrMetGluCysProAlaLeuIleValHisPro-------------------- 108
                  |||  :::   :::
Db   6128  CCCTCACCAGCGGCCGCTGCGCACCCTCTTCGCCACCCTTCGCCTGTCTGCCCTTCG 6069

Qy    109  --------ProAlaGlyGlyMetAlaSerGlySerGln------------ 119
                  :::  |||||||||
Db   6068  GCTCCTCCAGCACCTGCGCGCCCCTCTTCTCCCGGCAGCTCCTCAGCATTGCGCCCAGAG 6009
```

SQ   Sequence 6567 BP; 1494 A; 1736 C; 2190 G; 1147 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   580   Length: 6567
Score: 169.00   Matches: 56
Percent Similarity: 38.7%   Conservative: 19
Best Local Similarity: 28.9%   Mismatches: 91
Query Match: 14.9%   Indels: 28
DB: 8   Gaps: 5

US-09-989-890-238 (1-212) x AB22978 (1-6567)

Qy   2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
           ||| ||| ||| ||| |||:::
Db   6416 CCGGAGTACGCGGCAACGCCA---------CCACCCACTGCTGCCAGCAGTAGTACA 6369

Qy   22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAlaAla 41
           ||| ::: ||| ::: ||||||||| :::::::
Db   6368 CCACCTACACCTCCGGCAACAGCTGCCGTGGGAAGCCGCTGCCCTCCTTCTCCTCT 6309

Qy   42 CysSerGlyAspProGlyCysGlySerGlyAlaGlyLeu----------ProSer 56
           ||| ||||||||||||| ||||:::
Db   6308 TCTTCTGCACCGCCTTCGGTGGCTCCTCTGCTGGAGTTTCCGTCGGAGGGTTGCTCTTCA 6249

Qy   57 AlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAla 76
           :::||||||||||:::::||| ||| ||| :::
Db   6248 GGTCTCCAGCTTCAGGNGTTCGGGTGTTCGAGGGTTCTCAGGGGTTTCCGGCACTTGCTCG 6189

Qy   77 ProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLeuLysProThrGlyProArgSer 96
           ||| ||| ||| |||::: ||| :::
Db   6188 CCTCCTTCACTCGGCCCCTTCTCCCTCTCCAGGCTGGACCTCACCACCTTCACCGCCAGCA 6129

Qy   97 ThrMetGluCysProProAlaLeuIleValHisPro--------------- 108
           |||||||||
Db   6128 CCCTCACCAGGCGCGCCTGCGCCACCCTCTTCGCCGCCAGGTTCCCTGTCTCGCCTTCG 6069

Qy   109 ----------ProAlaGlyGlyMetAlaSerGlySerGln------------ 119
                     ::: |||:::|||
Db   6068 GCTCCTCCAGCACCTGCGCGCCCTCTTCTCCGGGCAGTCCTCAGCATTGCCGCCAGAG 6009

Qy   120 ---ProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerlysAlaSerLeuCysIle 138
             ||||||||||||| :::|||
Db   6008 CCTCCCTTCTCCGCCCGGCACCTGCGCCGCCCTCTTCTCCAGGCAGCTCCTCAGCACTACCG 5949

Qy   139 ProThrArgGlyProProGlnProLeu---MetArgThrProAlaAlaArgSerHis 157
           ||||||||||||| ||| :::||| ::: |||
Db   5948 CCAGAGCCTCCTCCTCCGCCACCTGCGCCCCTCTTCTCCGGGCAGCTCCTCAGCA 5889

Qy   158 TrpProIleProHisProCysAspThrAlaCysProAlaPro 171
           ||| |||:::||| |||:::::||||
Db   5888 CTACCGCCAGAGCCTCCAGCTTCGCCAGCACCCCCTTCTCCG 5847

Search completed: March 17, 2006, 04:05:23
Job time : 597 secs

THIS PAGE LEFT BLANK

OM protein - nucleic search, using frame_plus_p2n model

```
                    GenCore version 5.1.7
        Copyright (c) 1993 - 2006 Bioacceleration Ltd.
```

Run on: March 17, 2006, 04:05:36 ; Search time 3873 Seconds
                                          (without alignments)
                                          2561.028 Million cell updates/sec

Title: US-09-989-890-238
Perfect score: 1134
Sequence: 1 SPHQAAAPVDQTPRTLATMG........RRWAVAPCRABKLMCSSSRS 212

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop 6.0 , Fgapext 7.0
                Delop   6.0 , Delext 7.0

Searched: 41078325 seqs, 2339541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095247_16598/app_query.fasta_1
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss04
-USER=US09989890 @CGN_1_1_5315 @runat_16032006_095247_16598 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
              1: gb_est1:*
              2: gb_est2:*
              3: gb_est3:*
              4: gb_htc:*
              5: gb_est4:*
              6: gb_est5:*
              7: gb_est6:*
              8: gb_est7:*
              9: gb_est1:*
             10: gb_gss2:*
             11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Match Length	DB	ID	Description
1	1134	100.0	1197	11	DQ045548	DQ045548 Homo sapi
2	1032	91.0	1181	11	DQ045549	DQ045549 Pan trogl
3	1025	90.4	966	2	BE901537	BE901537 601677420
4	1024	90.3	917	5	BQ691555	BQ691555 AGENCOURT
5	985	86.9	1015	4	BE741035	BE741035 601594018
6	882.5	77.8	1671	4	CR749558	CR749558 Homo sapi
7	831	73.3	753	7	CN289134	CN289134 170005999
8	806	71.1	510	1	AA316608	AA316608 EST188290
9	790	69.7	583	3	BP331524	BP331524
10 c	687	60.4	623	2	BE741110	BE741110 601593919
11	685	60.4	685	11	AK051860	AK051860 Mus muscu
12	672	59.3	1009	5	BY707606	BY707606 Mus muscu
13	672	59.3	1657	4	AK007551	AK007551 Mus muscu
14	582	51.3	783	7	CK478920	CK478920 AGENCOURT
15	580	51.1	1040	3	BI411303	BI411303 602964692
16	554	48.9	680	3	BI653517	BI653517 603300520
17	516	45.5	825	8	DN936585	DN936585 AGENCOURT
18	507	44.7	582	3	BP315176	BP315176
19	482.5	42.5	944	3	BI905189	BI905189 603167516
20	479.5	42.3	746	3	BI905713	BI905713 603166180
21	468	41.3	826	3	BI694617	BI694617 603347838
22	462.5	40.8	756	3	BG871324	BG871324 602792482
23	454	40.0	425	2	BF150866	BF150866 uy88g10.y
24	454	40.0	1088	2	BG335025	BG335025 602403470
25	453	39.9	609	3	BE533148	BE533148 601235445
26	444	39.2	1092	2	BE409561	BE409561 601299865
27	427.5	37.7	426	1	AW762557	AW762557 ur57f12.y
28	427	37.7	769	7	CK473171	CK473171 AGENCOURT
29	414.5	36.6	722	2	BI409054	BI409054 602961119
30	393	34.7	244	2	BE162756	BE162756 PM1-HT045
31	383.5	33.8	738	2	BI078885	BI078885 602873115
32	368.5	32.5	928	2	BG335647	BG335647 602404274
33	366	32.3	523	1	AI789479	AI789479 uk99g01.y
34	363	32.0	485	2	BF041373	BF041373 BP250025B
35	342	30.2	402	6	CB769571	CB769571 AMGNNUC:M
36	317.5	28.0	798	2	BI106311	BI106311 602890634
37 c	281	24.8	616	2	BB466999	BB466999 BB466999
38	280	24.7	657	5	BY753126	BY753126 BY753126
39	279	24.6	685	2	BI156000	BI156000 602903535
40	268	23.6	746	2	BI159562	BI159562 602919525
41	262	23.1	537	2	BP191474	BP191474 239074 MA
42	245	21.6	419	6	CB764922	CB764922 AMGNNUC:N
43	242	21.3	535	5	BQ566109	BQ566109 gi52a07.y
44	239	21.1	139	1	AW732798	AW732798 bb14f07.Y
45	233	20.5	975	3	BM460277	BM460277 AGENCOURT

ALIGNMENTS

RESULT 1
DQ045548
LOCUS   DQ045548   1197 bp   DNA   linear   GSS 02-JUN-2005
DEFINITION Homo sapiens FLJ34633 gene, VIRTUAL TRANSCRIPT, partial sequence.
            genomic survey sequence.
ACCESSION   DQ045548
VERSION   DQ045548.1 GI:6896763
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 1197)
  AUTHORS   Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
            Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civello,D.,
            White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
  TITLE   A Scan for Positively Selected Genes in the Genomes of Humans and
            Chimpanzees
  JOURNAL   PLoS Biol. 3 (6), E170 (2005)
   PUBMED   15869325
REFERENCE   2 (bases 1 to 1197)
  AUTHORS   Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
            Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civello,D.,
            White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-MAY-2005) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
            them based on alignment. Translation starts at the beginning of
            alignment.

(Left column — Homo sapiens)

```
FEATURES             Location/Qualifiers
     source          1..1197
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
     gene            <1..>1197
                     /gene="FLJ34633"
                     /locus_tag="HC11301"
ORIGIN
```

Alignment Scores:
Pred. No.:
Score: 1.69e-56   Length: 1197
Percent Similarity: 1134.00   Matches: 212
Best Local Similarity: 100.0%   Conservative: 0
Query Match: 100.0%   Mismatches: 0
DB: 100.0%   Indels: 0
                                              Gaps: 0

US-09-989-890-238 (1-212) x DQ045548 (1-1197)

```
Qy    1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
Db  104 AGCCCCCACCAAGCCGCCGCCACCCGTAGACCACCAGACCCCAAGGACCCTGGCCACCATGGGC 163

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerArgProLeuProProAla  40
Db  164 CAGAGAGCATTACCTTCATCTCTGGCTCTCTGGCTCTGTCGAGTCCCCACCTGCT 223

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
Db  224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGGTGGTCCCGGGCTGCCTTCGCTTCCGCCGCT 283

Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysProLeu  80
Db  284 GCCGGGGATTGCCTTCCAGCCGCTGTGTGGCGGGATGCAGCCCCTGCGTCTA 343

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
Db  344 CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC 403

Qy  101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro 120
Db  404 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGNNNGGATGGCCAGCGGCTCAAGTCAACCA 463

Qy  121 TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
Db  464 TGGGCAGCAGTTCAGCTACCACCCCGATGTTAAGNTCAAAGGCATCCCTGTATCCCTACC 523

Qy  141 ArgGlyProProGlnProLeuMetArgThrArgProAlaAlaAlaArgSerHisTrpProIle 160
Db  524 CGAGGGGACCACCTCAGCCACTCCGCTGCCAAGGACCGCCGCCGCCGGCGCCGGATC 583

Qy  161 ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg 180
Db  584 CCCCACCCCATGCGACACAGCCTGCCCAGCACCTTGCCCGTAGTCCTGGTGGCCCGAGG 643

Qy  181 SerThrIleLeuSerMetSerArgThrThrCysArgArgTrpAlaValAlaAlaProCys 200
Db  644 AGTACTATTCTTTCCATGAGTCGGACCTGCCGGAGATGGGCAGTGGCTCCATGT 703

Qy  201 ArgAlaGluLysSerLeuMetCysSerSerSerArgSer 212
Db  704 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAGAGC 739
```

RESULT 2
DQ045549

```
LOCUS       DQ045549                1181 bp    DNA     linear   GSS 02-JUN-2005
DEFINITION  Pan troglodytes FLJ34633 gene, VIRTUAL TRANSCRIPT, partial
            sequence, genomic survey sequence.
ACCESSION   DQ045549
VERSION     DQ045549.1  GI:66896764
KEYWORDS    GSS.
```

(Right column — Pan troglodytes)

```
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment. Translation starts at the beginning of
            alignment.
FEATURES             Location/Qualifiers
     source          1..1181
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene            <1..>1181
                     /gene="FLJ34633"
                     /locus_tag="HC11301"
ORIGIN
```

Alignment Scores:
Pred. No.:
Score: 1.51e-50   Length: 1181
Percent Similarity: 1032.00   Matches: 195
Best Local Similarity: 92.0%   Conservative: 0
Query Match: 92.0%   Mismatches: 17
DB: 91.0%   Indels: 0
                                              Gaps: 0

US-09-989-890-238 (1-212) x DQ045549 (1-1181)

```
Qy    1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
Db  104 AGCCCCCACCAAGCCGCCGCCACCCGTAGACCACCAGACCCCAAGGACCCTGGCCACCATGGGC 163

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerArgProLeuProProAla  40
Db  164 CAGAGAGCATTACCTTCATCTCTGGCTCTCTGGCTCTGTCGAGTCCCCACCTGCT 223

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
Db  224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGGTGGTCCCGGGCNNNCTTNNGCTTCCGCCGCT 283

Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysProLeu  80
Db  284 GCCGGGGATNNNCCTCCAGCCGCTGTGTGGCGGGATGCAGCCCCTGCGTCTA 343

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
Db  344 CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC 403

Qy  101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro 120
Db  404 CCCCNNNNCCTGATCGTGCACCCCCCAGCCCCGNNNGGATGGCCAGCGNCTCAAGTCAACCA 463

Qy  121 TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
Db  464 TGGGCAGCAGTTCAGCTACCACCCCGATGTTAAGNTCAAAGGCATCCCTGTATCCCTACC 523

Qy  141 ArgGlyProProGlnProLeuMetArgThrArgProAlaAlaAlaArgSerHisTrpProIle 160
Db  524 ...
```

Db   524 CGAGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTGCGCAAGGAGCCACTGCCGATC 583

Qy   161 ProHisProCysAspThrAlaCysSerProAlaProLeuProValValLeuValAlaProArg 180
              |||||||||||||||||||||| |||||||||||||||||||||| || |||||||
Db   584 CCCCACCCCATGCGACCACAGCCTGCCCAGCACCCTTGCCCAGTAGTCCTCNTNGGCTNNGAGG 643

Qy   181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   644 AGTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCGCGGAGATGGGCAGTGGCTCCATGT 703

Qy   201 ArgAlaGluLysLeuMetCysSerSerSerArgSer 212
              ||||||||||||||||||||||||||||||||||
Db   704 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 739

RESULT 3
BE901537
LOCUS   BE901537   966 bp   mRNA   linear   EST 29-SEP-2000
DEFINITION   601677420F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3960220 5',
              mRNA sequence.
ACCESSION   BE901537
VERSION   BE901537.1   GI:10390818
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominidae; Homo.
REFERENCE   1 (bases 1 to 966)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
              Plate: LLCM842 row: g column: 05
              High quality sequence stop: 668.
FEATURES   Location/Qualifiers
     source   1..966
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:3960220"
              /tissue_type="choriocarcinoma"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_21"
              /note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
              Site_2: EcoRI; cDNA made by oligo-dT priming.
              Directionally cloned into EcoRI/XhoI sites using the
              following 5' adaptor: GGCACGAG(G). Size-selected >500bp
              for average insert size 1.8kb. Library constructed by
              Ling Hong in the laboratory of Gerald M. Rubin (University
              of California, Berkeley) using ZAP-cDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN
Alignment Scores:
Pred. No.:   3e-50   Length: 966
Score: 1025.00   Matches: 208
Percent Similarity: 98.1%   Conservative: 0
Best Local Similarity: 98.1%   Mismatches: 4
Query Match: 90.4%   Indels: 4
DB: 2   Gaps: 0

US-09-989-890-238 (1-212) x BE901537 (1-966)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   10 AGCCCCCACCAAGCGCCGCCACCCGTAGACCAGACCCCAAGGACCCCTGGC-ACCATGGGC 68

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGACCCTTGAGTCCCCACCTGCT 128

Qy   41 AlaCysSerGlyAlaAspProGlyCysGlySerGlyValGlyAlaGlyLeuProSerAlaSerAlaLeuAla 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   129 GCCTGCTCTGGCGACCTGGGTGTGGGAGTGGTGGCCGGCTGCTTTCTGCTTCCGCCGCT 188

Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   189 GCCGGGATTGCTTCCAGCGCGTGTGGAGCCTGTGTGGGGGATGCAGCCCCTGCCTGCTA 248

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   249 CTGACGACTCCACTGAGGGGACTGCTGAAGCCACCTGGCAAGGAGCACAATGAGTGC 308

Qy   101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnSerGlnPro 120
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   309 CCCCACCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAGCGGCTCAAGTCAACCA 368

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   369 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC 428

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIle 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   429 CGAGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTGCCAAGGAGCCACTGGCCGATC 488

Qy   161 ProHisProCysAspThrAlaCysSerProAlaProLeuProValValLeuValAlaProArg 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   489 CCCCACCCCATGCGACCACAGC-TGCCCAGCACCTTTGCCCAGTAGTCCTCGTGGCCTCCGAGG 547

Qy   181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys 200
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   548 AGTACTATTCTTTCCATGAGTCGGACCTGGAC-TGCCGGAGATGGGCAGTGGT-CCATGT 605

Qy   201 ArgAlaGluLysLeuMetCysSerSerSerArgSer 212
               ||||||||||||||||||||||||||||||||||
Db   606 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 641

RESULT 4
BQ691555
LOCUS   BQ691555   917 bp   mRNA   linear   EST 15-JUL-2002
DEFINITION   AGENCOURT_8341018 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6249453
              5', mRNA sequence.
ACCESSION   BQ691555
VERSION   BQ691555.1   GI:2181871
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominidae; Homo.
REFERENCE   1 (bases 1 to 917)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM2390 row: o column: 22
              High quality sequence stop: 535.
FEATURES   Location/Qualifiers
     source   1..917
              /organism="Homo sapiens"
              /mol_type="mRNA"

ACCESSION   BE741035
VERSION   BE741035.1   GI:1015027
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 1015)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished (1999)
JOURNAL
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM810 row: d column: 06
            High quality sequence stop: 752.
FEATURES   Location/Qualifiers
     source   1..1015
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3947861"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_9"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 1015
Score: 6.94e-48   Matches: 205
Percent Similarity: 96.7%   Conservative: 0
Best Local Similarity: 96.7%   Mismatches: 7
Query Match: 86.9%   Indels: 5
DB: 2   Gaps: 0

US-09-989-890-238 (1-212) x BE741035 (1-1015)

Qy   1 SerProHisGlnAlaAlaAlaAlaProValAspProArgThrLeuAlaThrMetGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50 AGCCCCCACCAAGCCGCGCAACCGTA-GACCAGACCCCAAGGACCCTGGC-ACCATGGGC 107

Qy   21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108 CAGAGAGCATTACCTTCATCTCTGCTCTCGGAGCCGGCCTTG-AGTCCCCCACCTGCT 166

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 167 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCCTGCCTTCTGCTTCCGCCGCT 226

Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 227 GCCGGGATTGCCTCCAGCGCTGTGGAGCCTGTGTGTGGGGATGCAGCCCCTGCCTGTCTA 286

Qy   81 LeuArgThrProLeuLeuLeuLysProThrGlyLeuLeuLysProThrMetGluCys 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 287 CTGAGGATTCCACTGAGGGGACTCGGCAAGGAGCCAAGGAGCACAATGGAGTGC 346

Qy 101 ProProAlaLeuIleValHisProProAlaGlyMetAlaSerGlySerGlnSerGlnPro 120

RESULT 5
LOCUS   BE741035   1015 bp   mRNA   linear   EST 15-SEP-2000
DEFINITION 601594018F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3947861 5',

```
                    PDPKDPGHGPESITFISGSABPALESPTCLLRTPLRGLLKPTGPRSTMECPPALIVH
                    PPAGGMASGSSQPWAAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCD
                    TACPAPLPVVLVAPRSTILSMSRTWTCRRWAVAPCRABKLMCSSSRS"

ORIGIN

Alignment Scores:
Pred. No.:
Score:                    1.27e-41              Length:               1671
Percent Similarity:       882.50               Matches:              172
Best Local Similarity:    81.1%                Conservative:         0
Query Match:              81.1%                Mismatches:           1
DB:                       77.8%                Indels:               40
                                               Gaps:                 4

US-09-989-890-238  (1-212) x CR749558 (1-1671)

Qy       1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
               |||||||||||||||||||||||||||||||||||||||||||||||
Db     172  AGCCCCCACCAAGCCGCCGACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC  231

Qy      21  GlnArgAlaAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla  40
Db     232  CAGAGAGCATTACCTTCATCTCGGCTCTGCTGAGCCGCCCTTGAGTCCC----------  282

Qy      41  AlaCysCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
Db     282  --------------------------------------------------------------  282

Qy     282  ----------------------------------------------------------  282
Db      61  AlaGlyIleAlaSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80
Db     283  --------------------------------------------------CCACC-TGTCTA  293

Qy      81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
               ||||||||||||||||||||||||||||||||||||||||||||||
Db     294  CTGAGGACTCCACTGAGGGGACTGCTGAAGCCACTGGGCCAAGGAGCACAATGGAGTGC  353

Qy     101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro  120
Db     354  CCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAGCGGCTCAAGTCAACCA  413

Qy     121  TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
               ||||||||||||||||||||||||||||||||||||||||||||||
Db     414  TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC  473

Qy     141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
Db     474  CGAGGGCCACCTCCGCAGCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGATC  533

Qy     161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg  180
               ||||||||||||||||||||||||||||||||||||||||||||||
Db     534  CCCCACCCATGCGACACAGCCTGCCCAGCACCTTTGCCAGTGTCCTGGCTCCGGAGG  593

Qy     181  SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys  200
Db     594  AGTACTATTCTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGGAGTGGCTCCATGT  653

Qy     201  ArgAlaGluLysLeuMetCysSerSerSerArgSer  212
               |||||||||||||||||||||||||||
Db     654  CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC  689
```

```
RESULT 7
CN289134
LOCUS          CN289134                    753 bp    mRNA      linear    EST  16-MAY-2004
DEFINITION     1700059993S935 GRN_PREHBP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION      CN289134
VERSION        CN289134.1    GI:47305548
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
  ORGANISM     Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1  (bases 1 to 753)
  AUTHORS      Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
```

Left column (bottom section):

```
Db     347  CCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAGCGGCTCAAGTCAACCA  406
               ||||||||||||||||||||||||||||||||||||||||||||||
Qy     121  TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
Db     407  TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC  466

Qy     141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
Db     467  CGAGGGCCACCTCCGCAGCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGATC  526

Qy     161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg  180
               ||||||||||||||||||||||||||||||||||||||||||||||
Db     527  CCCCACCCATGCGACACAGCCTGCCA-GCACCTTTGCCAGTAGTCCTGGCTCCGGAGG  585

Qy     181  SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys  200
Db     586  AGTACTATTCTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGGAGG-GCTCCATGT  644

Qy     201  ArgAlaGluLysLeuMetCysSerSerSerArgSer  212
               |||||||||||||||||||||||||||
Db     645  CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC  680

RESULT 6
CR749558
LOCUS          CR749558                   1671 bp    mRNA      linear    HTC  19-AUG-2004
DEFINITION     Homo sapiens mRNA; cDNA DKFZp686B17277 (from clone DKFZp686B17277).
ACCESSION      CR749558
VERSION        CR749558.1    GI:51476784
KEYWORDS       HTC.
SOURCE         Homo sapiens (human)
  ORGANISM     Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1  (bases 1 to 1671)
  AUTHORS      Poustka,A., Albert,R., Moomayer,P., Schupp,I., Wellenreuther,R.,
               Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
               Wiemann,S.
  CONSRTM      The German cDNA Consortium
  TITLE        Direct Submission
  JOURNAL      Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
               Neuherberg, GERMANY
  COMMENT      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
               Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
               sequenced by DKFZ (German Cancer Research Center,
               Heidelberg/Germany) within the cDNA sequencing consortium of the
               German Genome Project. This clone (DKFZp686B17277) is available at
               the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in
               Berlin, Germany. Please contact RZPD for ordering:
               http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686B17277
               Further information about the clone and the sequencing project is
               available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1..1671
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686B17277"
                     /tissue_type="prostate"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
                     /note="hypothetical protein, differentially spliced"
     gene            1..1671
                     /gene="DKFZp686B17277"
     CDS             69..692
                     /gene="DKFZp686B17277"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH18355.1"
                     /db_xref="GI:51476785"
                     /translation="MPRRGHPRPASGPPRLGPWERPTBLCLETYDKPPQPPPSRRTRR
```

TITLE   Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
            Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
PUBMED   15146197
COMMENT   Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 753   Std Error: 0.00.
FEATURES   Location/Qualifiers
     source   1..753
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, DMSO-treated H9 cell
                     line"
                     /clone_lib="GRN_PREHEP"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from DMSO-treated hES cell line H9 (p22) maintained in
                     feeder-free conditions"
ORIGIN

Alignment Scores:
Pred. No.:   4.71e-39   Length: 753
Score: 831.00   Matches: 157
Percent Similarity: 99.4%   Conservative: 0
Best Local Similarity: 99.4%   Mismatches: 0
Query Match: 73.3%   Indels: 1
DB: Gaps: 0

US-09-989-890-238 (1-212) x CN289134 (1-753)

Qy   56 SerAlaSerAlaAlaAlaGlyIleAlaSerAlaValGluProValCysGlyAspAla 75
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11 TCTGCTTCCGCGCTGCCGCGGATTGCCTCCAGCGCCGTGGAGCCTGTGTGCGGGGATGCA 70

Qy   76 AlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArg 95
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   71 GCCCCTGCCTGTCTACTGAGGACTCCACTGAGGGGACTCGTGAAGCCAACTGGGCCAAGG 130

Qy   96 SerThrMetGluCysGlyProProAla-LeuIleValHisProProAlaGlyGlyMetAlaSe 115
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   131 AGCACAATGGAGTGCGGCCCCCCAGCTGCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAG 190

Qy   115 rGlySerSerGlnProTrpProAlaAlaAlaSerArgAlaThrProMetLeuSerSerLysAlaSe 135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191 CGGCTCAAGTCAACCATGGCAGCAGCTTCAGCTACCCGATGTTAAGCTCAAAGGCATC 250

Qy   135 rLeuCysIleProThrArgGlyProProGlnProLeuMetArgThrProAlaAlaAr 155
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251 CCTGTGTATCCCTACCCGAGGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCAG 310

Qy   155 gSerHisTrpProIleProHisProCysAspThrAlaCysProAlaProLeuProValVa 175
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 GAGCCACTGGCCGGATCCCCACCCCATGCCGACACGGCCTGCCCAGCACCTTTGCCGTAGT 370

Qy   175 lLeuValAlaProArgSerThrIleLeuSerMetSerArgThrThrCysArgArgTr 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371 CCTCGTGGCCTGGCCTCCGGAGTAGTACTATTCTTTCCATGAGTCGGACCTGCCGGAGATG 430

Qy   195 pAlaValAlaProCysArgAlaGluLysGlyLeuMetCysSerSerSerArgSer 212
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   431 GGCAGTGGCCTCCATGTGCGAGCCGAGAAATTGATGTGCTCATCTTCAAGAGAGC 482

RESULT 8
AA316608
LOCUS   AA316608   510 bp   mRNA   linear   EST 19-APR-1997
DEFINITION EST188290 HCC cell line (metastasis to liver in mouse) II Homo

ACCESSION   AA316608
VERSION   AA316608.1   GI:1968936
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 510)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE   Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
   PUBMED   7566098
COMMENT   Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1..510
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):113362"
                     /db_xref="taxon:9606"
                     /tissue_type="colon"
                     /cell_type="KM12SM"
                     /cell_line="KM12C(HCC)metastasis into mouse (liver)"
                     /clone_lib="HCC cell line (matastasis to liver in mouse)
                     II"
                     /note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

Alignment Scores:
Pred. No.:   8.28e-38   Length: 510
Score: 806.00   Matches: 148
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 71.1%   Indels: 0
DB: Gaps: 1

US-09-989-890-238 (1-212) x AA316608 (1-510)

Qy   65 SerSerAlaValGluProValCysGlyValAspAlaAlaProAlaCysLeuLeuArgThrPro 84
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 TCCAGCGCCGTGGAGCCTGTGTGCGGGGATGCAGCCCCTGCCTGTCTACTGAGGACTCCA 61

Qy   85 LeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCysProArgProAlaLeu 104
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db    62  CTGAGGGACTGCTGAAGCCAACTGGGCAACATGGAGTGCCCCCAGCCCTG  121       Db    87  AGCCCCCACCAAGCCGCCGACCCGTAGACCAGACCAGACCCTGGCCACCATGGGC  146
Qy   105  IleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnProTrpAlaAlaAla  124       Qy    21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla  40
           ||||||||||||||||||||||||||||||||||||||||||||||                            ||||||||||||||||||||||||||||||||||||||||||
Db   122  ATCGTGCACCCCCCAGCCGGCGGCGGATGGCCAGCGGCTCAAGTCAACCATGGGCCAGCT  181       Db   147  CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGGCCGGGTCCCCCACCTGCT  206
Qy   125  SerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProPro  144       Qy    41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
           ||||||||||||||||||||||||||||||||||||||||||||||                            ||||||||||||||||||||||||||||||||||||||||||
Db   182  TCAGCTACCCCGATGTAAGCTCAAAGGCATCCCTGTGTATCCCTACCCGAGGGCCACCT  241       Db   207  GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCCGGGCTGCCTTCTGCTTCCGCGCT  266
Qy   145  ProGlnProLeuProMetArgThrProAlaAlaArgSerHisTrpProIleProHisProCys  164       Qy    61  AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80
           ||||||||||||||||||||||||||||||||||||||||||||||                            ||||||||||||||||||||||||||||||||||||||||||
Db   242  CCCCAGCCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGATCCCCACCCATGC  301       Db   267  GCCGGGATTGCCTCCAGCGGTGGAGCCTGTGTGCGGGGATGCAGCGCCCCTGCCTGTCTA  326
Qy   165  AspThrAlaCysSerProLeuProLeuValValAlaAlaProAlaArgSerThrIleLeu  184       Qy    81  LeuArgThrProLeuArgGlyLeuLeuLeuProThrGlyThrGlySerThrMetGluCys  100
           ||||||||||||||||||||||||||||||||||||||||||||||                            ||||||||||||||||||||||||||||||||||||||||||
Db   302  GACACAGCCTGCCCAGCACCTTGCCAGTAGTCCTCGTGCGTCCGAGGAGTACTATTCTT  361       Db   327  CTGAGGACTCCACTGAGGCAGCTGCTGAAGCCAAGAGGCCAAGGACAACAATGGAGTGC  386
Qy   185  SerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCysArgAlaGluLys  204       Qy   101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPro  120
           ||||||||||||||||||||||||||||||||||||||||||||||                            ||||||||||||||||||||||||||||||||||||||||||
Db   362  TCCATGAGTCGGACCTGGACCTGCCGGAGAGTGGGCAGTGCCTCCATGTCGAGCCGAGAA  421       Db   387  CCCCAGCCCCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAGCGGCTCAAGTCAACCA  446
Qy   205  LeuMetCysSerSerSerArgSer  212                                              Qy   121  TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
           ||||||||||||||||||                                                          ||||||||||||||||||||||||||||||||||||||||||
Db   422  TTGATGTGCTCATCTTCAAGAAGC  445                                              Db   447  TGGGCAGCAGCTTCAGCTACCCCGATGTAAGCTCAAAGGCATCCCTGTGTATCCCTACC  506

RESULT 9                                                                          Qy   141  ArgGlyGlyProProGlnProLeuProMetArgThrProAlaAlaArgSerHisTrpProIle  160
BP331524                                                                                   ||||||||||||||||||||||||||||||||||||||||||
LOCUS      BP331524      583 bp    mRNA      linear   EST 17-SEP-2004         Db   507  CGNAGGCCAANCTTCCCAGCNCCGATGCGGACTCCTGCTGC-AGGGACCACTGGCCGAT-  564
DEFINITION BP331524 Sugano cDNA library, rectum Homo sapiens cDNA clone
           RCT09504, mRNA sequence.                                                Qy   161  ProHisProCysAspThr  166
ACCESSION  BP331524                                                                         ||||||||||||||||
VERSION    BP331524.1  GI:52260875                                              Db   565  CCCCACCATGCGACACA  582
KEYWORDS   EST.
SOURCE     Homo sapiens (human)                                               RESULT 10
  ORGANISM Homo sapiens                                                        BE741110
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   LOCUS      BE741110       623 bp    mRNA    linear   EST 15-SEP-2000
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;         DEFINITION 601593919F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3947885 5',
           Hominidae; Homo.                                                                mRNA sequence.
REFERENCE  1  (bases 1 to 583)                                                 ACCESSION  BE741110
  AUTHORS  Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,       VERSION    BE741110.1  GI:10155102
           Mizushima,S., Sugano,J., Nakai,K. and Sugano,S.                     KEYWORDS   EST.
  TITLE    Sequence comparison of human and mouse genes reveals a homologous   SOURCE     Homo sapiens (human)
           block structure in the promoter regions                              ORGANISM  Homo sapiens
  JOURNAL  Genome Res. 14 (9), 1711-1718 (2004)                                           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  PUBMED   15342556                                                                       Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
COMMENT    Contact: Yutaka Suzuki                                                         Hominidae; Homo.
           Department of Virology                                              REFERENCE  1  (bases 1 to 623)
           Institute of Medical Science, University of Tokyo                     AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
           4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan                  TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
           Email: ysuzuki@ims.u-tokyo.ac.jp.                                     JOURNAL  Unpublished (1999)
FEATURES            Location/Qualifiers                                        COMMENT    Contact: Robert Strausberg, Ph.D.
     source         1..583                                                                Email: cgapbs-r@mail.nih.gov
                    /organism="Homo sapiens"                                              Tissue Procurement: DCTD/DTP
                    /mol_type="mRNA"                                                       cDNA Library Preparation: Ling Hong/Rubin Laboratory
                    /db_xref="taxon:9606"                                                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                    /clone="RCT09504"                                                      DNA Sequencing by: Incyte Genomics, Inc.
                    /tissue_type="rectum"                                                  Clone distribution: MGC clone distribution information can be
                    /clone_lib="Sugano cDNA library, rectum"                              found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
ORIGIN                                                                                     Plate: LLCM810 row: e column: 06
                                                                                          High quality sequence stop: 615.
Alignment Scores:                                                             FEATURES            Location/Qualifiers
 Pred. No.:                                             Length:        583         source         1..623
 Score:              8.45e-37         790.00            Matches:       158                        /organism="Homo sapiens"
 Percent Similarity:  95.2%                             Conservative:    0                        /mol_type="mRNA"
 Best Local Similarity: 95.2%                           Mismatches:      8                        /db_xref="taxon:9606"
 Query Match:         69.7%                             Indels:          2                        /clone="IMAGE:3947885"
 DB:                  3                                 Gaps:            0                        /tissue_type="adenocarcinoma cell line"
                                                                                                 /lab_host="DH10B (phage-resistant)"
US-09-989-890-238 (1-212) x BP331524 (1-583)                                                      /clone_lib="NIH_MGC_9"
                                                                                                 /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
Qy     1  SerProHisGlnAlaAlaAlaProValAlaProThrProArgThrLeuAlaThrMetGly  20                         EcoRI; cDNA made by oligo-dT priming. Directionally
           |||||||||||||||||||||||||||||||||||||||||||||                                           cloned into EcoRI/XhoI sites using the following 5'
```

ORIGIN

adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

Alignment Scores:

Pred. No.:		
Score:	9.55e-31	Length:	623
Percent Similarity:	687.00	Matches:	164
Best Local Similarity:	84.4%	Conservative:	4
Query Match:	82.4%	Mismatches:	17
DB:	60.6%	Indels:	14
	2	Gaps:	2

US-09-989-890-238 (1-212) x BE741110 (1-623)

```
Qy     1 SerProHisGlnAlaAlaAla-ProValAsp-GlnThrProArgThrLeuAlaThrMetG  20
           ||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    51 AGCCCCCACCAGCCGCCGCCCACCCAGTAGACCATGACCCCAAGGACCCTGGCCACCATCG 110

Qy    20 lyGlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProA  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   111 GCCAGAGAGCCATTACCTTCATCTCTGGCTCTGCTGAGCCGGCCCCTTGAGTCCCCACCTG 170

Qy    40 laAlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla-  59
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   171 CTGCCTGCTCCTGGCGACCCTGGGTCGGGAGTGGTGCCGGCCTGCCTTCGCGCC 230

Qy    60 AlaAlaGlyIleAlaSerSer-AlaValGluProValCysGlyAspAlaAlaProAlaCy  79
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   231 GCTGCCGGGATTGCCTCCAGTCGCTGTGAGCCTGTGTGTGGGGATGCAGCCCCCTGCCTG 290

Qy    79 aLeuLeuArgThrProLeuArgGlyLeuLeuLysProThrGly--ProArgSer---ThrM  98
           ||||||||||||||||||||||||||||||||||||||||
Db   291 TCTACTACGGGATTCCACTGAGGGACTCTGAGCCCACTGGTGCCCAGGAGTCACAATG 350

Qy    98 etGluCysProProAlaLeuIle-ValHisProPro-AlaGlyMetAlaSerGly-S  117
           |||||||||||||||||||||||||||||||||||||||||| :
Db   351 GTAGTGACCCCCCAGCCCTGATTCGTGCACCCCCCAGACCCGGGATGGCCAGCGGCTG 410

Qy   117 erSerGln-ProTrpAlaAlaAla---SerAlaThrProMetLeuSerSerLysAlaSer 135
           ::|||:|| ||| ||: |||: ||||| |||||
Db   411 CAAGTCAATCCATGGGACAGGCAGGTCTTCAGCTATCCCCGATGTTAAGCTCAAAGCCATCC 470

Qy   136 LeuCysIleProThrArg-GlyProProGlnProLeuMetArgProLeuProAla-AlaA 155
           |||||||||||||||||||||||||||||||||||||||||||||
Db   471 CTGTGTATCCCCTACCCGAGGGTCCACCTTCCCCAGCCCTGATGCGGACTCCTGCTGACAA 530

Qy   155 rgSerHisTrpProIleProHisProCysAsp-ThrAlaCysProAlaProLeuProVal 174
           || :|| ||| ||| :|:: ||| ||| |||
Db   531 GGAGCCACTGGCCGATCTCCCAACCATGCGAGCACAGGTGACCAGTACCTTGGCCAGTA 590

Qy   175 ValLeuValAlaProArgSerThrIleLeuSer 185
           |||||||||||||||||
Db   591 GTCCTGCTGGCTTCCGAGGAGTACTATACTTTCC 623
```

RESULT 11

AK051860/c

LOCUS   AK051860   1596 bp   mRNA   linear   HTC 03-APR-2004

DEFINITION   Mus musculus 12 days embryo eyeball cDNA, RIKEN full-length
    enriched library, clone:D230011M17 product:hypothetical protein,
    full insert sequence.

ACCESSION   AK051860

VERSION   AK051860.1   GI:26342255

KEYWORDS   HTC; CAP trapper.

SOURCE   Mus musculus (house mouse)

ORGANISM   Mus musculus

    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
    Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1

AUTHORS   Carninci,P. and Hayashizaki,Y.

TITLE   High-efficiency full-length cDNA cloning

JOURNAL   Meth. Enzymol. 303, 19-44 (1999)

PUBMED   10349636

REFERENCE   2

AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
    Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
    prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)

PUBMED   11042159

REFERENCE   3

AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
    Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
    Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
    Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
    Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
    Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
    Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE   RIKEN integrated sequence analysis (RISA) system-384-format
    sequencing pipeline with 384 multicapillary sequencer

JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)

PUBMED   11076861

REFERENCE   4

AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
    FANTOM Consortium.

TITLE   Functional annotation of a full-length mouse cDNA collection

JOURNAL   Nature 409, 685-690 (2001)

REFERENCE   5

AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
    Group Phase I & II Team.

TITLE   Analysis of the mouse transcriptome based on functional annotation
    of 60,770 full-length cDNAs

JOURNAL   Nature 420, 563-573 (2002)

REFERENCE   6 (bases 1 to 1596)

AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
    Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
    Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
    Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
    Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
    Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,S., Murata,M.,
    Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
    Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
    Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
    Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
    Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
    Muramatsu,M. and Hayashizaki,Y.

TITLE   Direct Submission

JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
    Physical and Chemical Research (RIKEN), Laboratory for Genome
    Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
    RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
    Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
    URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
    Fax:81-45-503-9216)

COMMENT   cDNA library was prepared and sequenced in Mouse Genome
    Encyclopedia Project of Genome Exploration Research Group in Riken
    Genomic Sciences Center and Genome Science Laboratory in RIKEN.
    Division of Experimental Animal Research in Riken contributed to
    prepare mouse tissues.
    Please visit our web site for further details.
    URL:http://genome.gsc.riken.jp/.
    URL:http://fantom.gsc.riken.jp/.

FEATURES   Location/Qualifiers

source   1..1596
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6J"
    /db_xref="FANTOM_DB:D230011M17"
    /db_xref="taxon:10090"
    /clone="D230011M17"
    /tissue_type="eyeball"
    /clone_lib="RIKEN full-length enriched mouse cDNA library"
    /dev_stage="12 days embryo"

```
CDS              890..1432
                 /note="unnamed protein product; hypothetical protein
                 (evidence: decoder,Longest-ORF)
                 putative"
                 /codon_start=1
                 /protein_id="BAC34790.1"
                 /db_xref="GI:26342256"
                 /translation="MLVLSLWPSRRLGGARSGLGGTPLCSLAQAASADPSMGSPADKQ
                 GLQPRTQAPQRCRQSLQRRKQKAALHQSPQPGQRRQVGGSLAGSAEPEMKVILSGP
                 WNPGSLGSGLLVRRPGGGWERSSFVSRHNSDGLSQGPKRGGPDDGRGCPGLGMAQSMA
                 CCSQAPAGRSQBEVRDKLLG"
ORIGIN
Alignment Scores:
Pred. No.:        1596
Score:            685.00        Length:         1596
Percent Similarity: 71.2%       Matches:        143
Best Local Similarity: 67.5%    Conservative:   8
Query Match:      60.4%         Mismatches:     61
DB:                4            Indels:         0
                                Gaps:           0

US-09-989-890-238 (1-212) x AK051860 (1-1596)
```

```
Qy     1  SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
          ||||||||||||::|||  ||||||    |||||  ||||||||||||||||||||||||
Db  1250  AGCCCCCACCGGGCCTCGCACCAGGAGGCCAGACCCTAAGGACCCTGGCCACCACGGGC  1191

Qy    21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla  40
          ||||||||||||||||||||||||||       ::|||||||||||||||||||||||
Db  1190  CAGAGAGTATCACCTTCATTTCAGGCTCTGCAGAACCAGCCAACCAGCCCCAACCTGCT  1131

Qy    41  AlaCysSerGlyValAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
          |||  |||||||||||||||||||||||||||::||||||||||||||||||
Db  1130  GCCTCCTCTGGCGCCCTGGGGTTGGACTGGTAGGGCTGCCTTCTGCTTCCGACGCT  1071

Qy    61  AlaGlyIleAlaSerArgAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80
          |||||||||||||||||||||||||||||||||||||||  |||||||||||||||
Db  1070  GCAGGGGATTGCCTGCAGCCGTGGGGGCTTGTGTGCGGGCTGCGAGCCCCTGCTTGTCTG  1011

Qy    81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
          |||||||||||||||||||||||||||||||||::|||::||||||||||||||
Db  1010  CCGGAGACCCCATTGAAGGGTCTGCGAAGCCGCCTGGGCCAAGAACACAATGGTGTGC  951

Qy   101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPro  120
          |||||||||::||||| |||||||||||||||||||||||||||||||||||||||||
Db   950  CCCCCAGCCCAGACCGTGCACCCCCCAGCCCGCCGGGATGGCCAGAGGCTCAAGACAAGCA  891

Qy   121  TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
          ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db   890  TGGGCCAGCAGCTTCAGCTACCCTGATGTTAAGCTCAAAAGGCATCCTGTCTACCCCTACC  831

Qy   141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisProProIle  160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   830  GCCATGCCCACCTCCCCAGTCCTGACGTGGACTCCTGCTGCTGCAAGGAGCCCCTGGCCGAGC  771

Qy   161  ProHisProCysAspThrAlaCysSerProAlaProLeuProValValLeuValAlaProArg  180
          ||||||||||||||||||||||||||||||||||  ||| ||||||||||||||||
Db   770  CTCTCCCACGCGGCACAGCTTGCCTAGCACCTTCACCAACAGCCCCGGGCCTCGAGG  711

Qy   181  SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpProValAlaProCys  200
          |||||  |||::|||||||||||||||||||||||||||||||||||||||||||||
Db   710  AGTACTACTCCTTCCATGAATCGGACCTGCCTGACATGGGGCAGTGGCTCCATGT  651

Qy   201  ArgAlaGluLysLeuMetCysSerSerSerArgSer  212
          |||  ::|||||||||||||||||||||||||||||
Db   650  CGAGGCGGGAGATCGACGTGCTTATTTCAAGAAGC  615
```

```
RESULT 12
BY707606
LOCUS       BY707606     1009 bp    mRNA      linear    EST 16-DEC-2002
DEFINITION  BY707606 RIKEN full-length enriched, 10 day old male pancreas Mus
            musculus cDNA clone 1810019J16 5', mRNA sequence.
ACCESSION   BY707606
```

```
VERSION    BY707606.1  GI:27118783
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
           Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE  1  (bases 1 to 1009)
AUTHORS    Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
           Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
           Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
           Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
           Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
           Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
           Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
           Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
           Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
           Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
           Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
           Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
           Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
           Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
           Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
           Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
           Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
           Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
           Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y.,
           Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
           Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
           Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
           Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
           Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
           Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
           Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
           Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE      Analysis of the mouse transcriptome based on functional annotation
           of 60,770 full-length cDNAs
JOURNAL    Nature 420, 563-573 (2002)
PUBMED     12466851
COMMENT    Contact: Yoshihide Hayashizaki
           Laboratory for Genome Exploration Research Group, RIKEN Genomic
           Sciences Center(GSC), Yokohama Institute
           The Institute of Physical and Chemical Research (RIKEN)
           1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           Tel: 81-45-503-9222
           Fax: 81-45-503-9216
           Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
           Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
           Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
           Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
           Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
           Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
           Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
           Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
           Direct Submission
           Computational Analysis of Full-Length Mouse cDNAs Compared with
           Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
           Normalization and subtraction of cap-trapper-selected cDNAs to
           prepare full-length cDNA libraries for rapid discovery of new
           genes. Genome Res. 10 (10), 1617-1630 (2000)
           RIKEN integrated sequence analysis (RISA) system--384-format
           sequencing pipeline with 384 multicapillary sequencer. Genome Res.
           10 (11), 1757-1771 (2000)
           Computer-based methods for the mouse full-length cDNA
           encyclopedia: real-time sequence clustering for construction of a
           nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
           cDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues.
           Please visit our web site (http://genome.gsc.riken.go.jp) for
           further details.
```

FEATURES   Location/Qualifiers
    source   1..1009
                  /organism="Mus musculus"
                  /mol_type="mRNA"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="1810019J16"
                  /sex="male"
                  /tissue_type="pancreas"
                  /dev_stage="10 day old"
                  /clone_lib="RIKEN full-length enriched, 10 day old male
                  pancreas"
ORIGIN

Alignment Scores:
Pred. No.:   1.33e-29   Length: 1009
Score: 672.00   Matches: 143
Percent Similarity: 70.9%   Conservative: 8
Best Local Similarity: 67.1%   Mismatches: 61
Query Match: 59.3%   Indels: 1
DB: 5   Gaps: 0

US-09-989-890-238 (1-212) x BY707606 (1-1009)

```
Qy     1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
           ||||||||||||:::|||   |||||||||||:||||||||||||||||
Db   147 AGCCCCACCGGGACCGTCGCACCAGGAGGCCAGACCCTAAGGACCCTGGCCACCACGGGC 206

Qy    21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
           |||||         |||||      ::::||| |||||| |||||| ||||
Db   207 CAGAGAGATCACCTTCATTTCAGGCTCTGCAGAACCAGCCAACGAGCCCCAACCTGCT 266

Qy    41 AlaCysSerGlyAlaAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
           ||||||||||||||  ||  ||||||||||||||
Db   267 GCCTCCTCTGGCGCCCCCTGGGGTTGGGACTGGTGTAGGGCTGCGTCCTTCTGCTTCCGACGCT 326

Qy    61 AlaGlyIleAlaSerArgAlaValGluProValCysGlyGlyAspAlaAlaProAla-CysLe 80
           ||||||||| |||||| ||||||| |||| |||||||||||||||||||
Db   327 GCAGGGATTGCCTGCAGCGCGTGTGGGGCTTGTGTCGGGGCTGCAGCCCCTGCATTGTCT 386

Qy    80 uLeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCy 100
           ||||||||:::|||| |||||||||||:::|||||||||| |||||||
Db   387 GCCGGAGACCCCATTGAAGGGTTCGCCGAAGCCGCCTGGGCCAAGGACACAATGGTGTG 446

Qy   100 sProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPr 120
           |||||||||||||||||||||||||||||  |||
Db   447 CCCCCCAGCCAGACCGTCACCCCCGCCGGGGATGGCCAGAGGCTCAAGACAAGC 506

Qy   120 oTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProTh 140
           |||||||||||| |||:|| |||||||:|| |||||||||||
Db   507 ATGGGCAGCAGCTTCAGCTACCCTGATGTTAAGCTCAAAGGCATCCCTGTCTACCCCTAC 566

Qy   140 rArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProProIl 160
           |:::|| |||||||||| ||||||| || |||||||||
Db   567 CGCCATGCCACCTCCCCAGTCCCTGACGTGGACTCCTGTCTGCAAGGAGCCCCCTGGCCAG 626

Qy   160 eProHisProCysAspThrAlaProLeuProValValLeuValAlaProAr 180
           ||||||||| || ||| |||||||||||:|| ||||||||||||
Db   627 CCTCCTCCCACGCGCGGCACAGCTTGCCTAGCACCTTCACCCAACAGCCCCCGGCGCTCGAG 686

Qy   180 gSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgArgProCy 200
           |||| ||:::||||:|| |||||||||||||||||
Db   687 GAGTACTACTCCTTCCATGAATCGGACCCTGCCTGAGATGGGCCAGTGGCTGCCCCATG 746

Qy   200 sArgAlaGluLysLeuMetCysSerSerSerArgSer 212
           ||:::             |||||||||
Db   747 TCGAGCCGGGAGATCGACCTGCTTATTTTCAAGAAGC 783
```

RESULT 13
AK007551
LOCUS   AK007551   1657 bp   mRNA   linear   HTC 03-APR-2004
DEFINITION Mus musculus 10 day old male pancreas cDNA, RIKEN full-length
            enriched library, clone:1810019J16 product:SIMILAR TO TATA BOX

ACCESSION   BINDING PROTEIN (TBP)-ASSOCIATED FACTOR, RNA POLYMERASE II, C1,
            130KD homolog [Mus musculus], full insert sequence.
            AK007551
VERSION   AK007551.1 GI:12841167
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,H., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 1657)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   Please visit our web site (http://genome.gsc.riken.jp/) for further
            details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer

```
         [5' GAGAGAGAAGAAGATCCAAGAGCTCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
         prepared by using trehalose thermo-activated reverse transcriptase
         and subsequently enriched for full-length by cap-trapper. cDNA went
         through one round of subtraction to Rot = 20.0. Second strand cDNA
         was prepared with the primer adapter of sequence[5'
         GAGAGAGAGAGCGGGCCCAATTAATTCTCGAGTAATTAAATTAATCCCCCCCCCC 3'], cDNA
         was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'
         end: SstI. Host: SOLR.
FEATURES          Location/Qualifiers
     source        1..1657
                   /organism="Mus musculus"
                   /mol_type="mRNA"
                   /strain="C57BL/6J"
                   /db_xref="FANTOM_DB:1810019J16"
                   /db_xref="taxon:10090"
                   /clone="1810019J16"
                   /sex="male"
                   /tissue_type="pancreas"
                   /clone_lib="RIKEN full-length enriched mouse cDNA library"
                   /dev_stage="10 day old"
     CDS           138..1241
                   /note="unnamed protein product; SIMILAR TO TATA BOX
                   BINDING PROTEIN (TBP) -ASSOCIATED FACTOR, RNA POLYMERASE
                   II, C1, 130KD homolog [Mus musculus] (SPTR|Q92200,
                   evidence: FASTY, 97.2%ID, 100%length, match=649)
                   putative"
                   /codon_start=1
                   /protein_id="BAC25177.1"
                   /db_xref="GI:26357449"
                   /translation="MSAPSPHRAVAPGGQTLRTLATTGQRVSPSFQALQNQPTSPQPA
                   ASSGAPGVGTVGLPSASDAAGIACSAVGLVCGAAAPALSAGDPIEGSABAAWAKEHN
                   GVPPSPDRAPPSRRDGQRLKTSMGSSFSYPDVKLKGIPVYPYRHATSPVPDVDSCCKE
                   PLAEBPPPRHSLPSTFTNSPRGSBEYYSFHESDLDLPEMGSGSMSSRRIDVLFFKKLT
                   ELFSVHQIDELAKCTSDTVFLEKTSKISDLISSITQDYHLDEQDAEGRLVRGIIRIST
                   RKSRSRPQTSEGRSARSTAPAAAPDSGHETMLGSGLSQDELTVQISQETTADAIARKL
                   RPYGAPGYPASQDSSFQGTDTDSSGAPLLQVYC"

ORIGIN

Alignment Scores:
Pred. No.:                2.5e-29        Length:           1657
Score:                    672.00         Matches:          143
Percent Similarity:       70.9%          Conservative:     8
Best Local Similarity:    67.1%          Mismatches:       61
Query Match:              59.3%          Indels:           1
DB:                       4              Gaps:             0

US-09-989-890-238 (1-212) x AK007551 (1-1657)

Qy      1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
          |||||||||||:::|||    ||||||||||||||||||||||||||||||||||||
Db    150 AGCCCCCACGCCGGGCCGTCGGACCACCAGGAGGCCAGAACCCTAAGGACCCTGCCACCACGGGC  209

Qy     21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuLeuSerArgProLeuSerProLeuProProAla  40
          |||||    ||  |    :::|||||||||||||||||||||||||||||||||||
Db    210 CAGAGAGTATCACCTTCATTTCAGGGCTCTGCAGAACCAGCCAACGAGCCCCCAACCTGCT  269

Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerArgAlaSerAlaAla  60
          |||    ||||||||||||||||:::|||||||||||||||||||||||||||||||
Db    270 GCCTCCTCTGGCGCCCTGGGGTTGGGACTGGTGTGTAGGGCTGCCTTCGCTTCCGACGCT  329

Qy     61 AlaGlyIleAlaSerSerAlaValAlaGluProValCysGlyAspAlaAlaProAla-CysLeu  80
          |||||||||||||||    |    |||||||    ||||||    ||||||||||||||
Db    330 GCAGGGATTGCCTCAGCCGCTGTGGGCGTTGTGCGGGCTGCAGCCCCTGCATTGTCT  389

Qy     80 uLeuArgThrProLeuArgGlyLeuLeuLysProThrGlySerThrMetGluCys  100
          |||||||||    |||    :::||||||||||||||||||||||||||||||||
Db    390 GCCGGAGACCCCATTGAAGGGTCTGCCGAAGGCCGCTGGGCCAAGAACAACAATGGTGTG  449

Qy    100 sProProAlaLeuIleValHisProProAlaGlyValGlyMetAlaSerGlySerGlnPr  120
          |||||||||||||    |||||||||||||||||||||||||||||||||||
Db    450 CCCCCAGCCAGGACCGTGCAYCCCCAGCCAGGGGATGGCCAGAAGCTCAAGACAAGC  509

Qy    120 oTrpAlaAlaAlaThrProMetLeuSerLeuSerLysAlaSerLeuCysIleProTh  140
```

```
Db    510 ATGGGCAGCAGCTTCAGCTACCTACCCGATGTTAAGCTCAAAGCATCCTGTCTACCCTAC  569
          ||||||||||||||||||||||||||||||||||||||||||||||    ||||
Qy    140 rArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIl  160
          ||||||||||||||||||||    ||||||||||||||||||:::|||||
Db    570 CGCCATGGCCACCTCCCCAGTCCTGAGCGTGGACTCCTGCTGCAAGGAGCCCCTGGCCGAG  629
          |||||||||||    |||||    |||||||||||||||||
Qy    160 eProHisProCysAspThrArgAlaCysProArgProLeuProValValLeuValAlaProAr  180
          |||    :::||||||    ||||||||||||||||||||||||
Db    630 CCTCCTCCACGCGGCACAGCTTGCCTAGCACCTTCACCAACAGCCCCCGGGGCTCTGAG  689
          |||||||||||||||||||||||||||||||||||||||||||
Qy    180 gSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCy  200
          ||||||    |||||||:::|||||||||||||||||||||
Db    690 GAGTACTACTCCTTCCATGATGACCTGCCTGAGATGGGCAGTGGCTCCATG  749
          ||||||||||||||||||||||||||||||||||||||||||
Qy    200 sArgAlaGluLysMetCysSerSerArgSer  212
          |||       :::
Db    750 TCGGAGCCGGGAGATCCAGCGTGCTTATTTTCAAGAAGC  786
```

```
RESULT 14
CK478920
LOCUS       CK478920                     783 bp    mRNA       linear    EST 14-JAN-2004
DEFINITION  AGENCOURT_17582913 NIH_MGC_232 Rattus norvegicus cDNA clone
            IMAGE:7121341 5', mRNA sequence.
ACCESSION   CK478920
VERSION     CK478920.1  GI:40823054
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 783)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Howard Jacobs
            cDNA Library Preparation: Express Genomics
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM15005 row: p  column: 11
            High quality sequence stop: 736.
FEATURES          Location/Qualifiers
     source        1..783
                   /organism="Rattus norvegicus"
                   /mol_type="mRNA"
                   /db_xref="taxon:10116"
                   /clone="IMAGE:7121341"
                   /tissue_type="lung, pooled"
                   /lab_host="DH10B TonA"
                   /clone_lib="NIH_MGC_232"
                   /note="Organ: lung; Vector: pExpress-1; Site 1: EcoRV;
                   Site 2: NotI; RNA obtained from pooled lung tissue from a
                   mix of male and female animals at 8 wk old. Tissues were
                   snap-frozen and kept at -80C for two days before RNA
                   extraction and purification (Tri-reagent method). cDNA was
                   primed using oligo-dT primer:
                   5'-PGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                   the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb
                   resulted in an average insert size of 2.3 kb. This primary
                   library is normalized (non-normalized primary library is
                   NIH_MGC 231) and was constructed by Express Genomics
                   (Frederick, MD). Note: this is a NIH_MGC library."
ORIGIN
```

```
Alignment Scores:
Pred. No.:                              Length:               783
Score:                   1.74e-24       Matches:              582.00
Percent Similarity:      68.4%          Conservative:         125
Best Local Similarity:   64.8%          Mismatches:           7
Query Match:             51.3%          Indels:               61
DB:                                     Gaps:                 0
                                                              0

US-09-989-890-238 (1-212) x CK478920 (1-783)

Qy      1  SerProHisGlnAlaAlaAlaProValAspProArgThrLeuAlaThrMetGly  20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194  AGCCCCCACCAGGGCGGTCGCACCCGGAGGCCAGACCCTAAGGACCCTGGCCACCACGGGGC  253

Qy     21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla  40
           |||||||||||||||||||||||||    ::|||||||
Db    254  CAGAGAGCATCACCTTCATTCAGGCTCTGCAGAACCAACAGCCCCAACCTGCT  313

Qy     41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
           ||||||||||||||||:::|||||:::||||||||||||||||
Db    314  GCCTCCTCTGGGCGCCCCTGGGGTTGGACTGGTGTAGGGCTGCCTTCTTGCTTCCGACGCT  373

Qy     61  AlaGlyIleAlaSerSerAlaValAlaGluProValCysGlyAspAlaAlaProAlaCysLeu  80
           ||||||||||||||||||||||:::|||||||:::||
Db    374  GCAGGGAATTGCCTCAGCCGCTGGAGCTTGTGTGCGAGCTGCCCCCTGCTTATCTG  433

Qy     81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
           ||||||:::||||||||||::::|||||||||||:::||
Db    434  CTGGAGACCCCATTGAAGGGTCTTCGGAAGCCCGGCCCAAGGAACACAACGGTGTGC  493

Qy    101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlnSerSerGlnPro  120
           ||||||||||||||||||||||||||||||||::||
Db    494  CTCCCAGCCCAGACCGAGCACCCCCCAGCCGCGCGGGATGGCCAAAAGTCAAGACCAGCA  553

Qy    121  TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
           |||||||||||||||||||||||:::|||||||||
Db    554  TGGGCAGCAGCTTCAGCTACCCTGATGTTAAGCTCAAGGGCATCCCGGTCTACCCCTACC  613

Qy    141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpThrCysArg  160
           ||||||||:::||||||||||||:::||
Db    614  GCCATGCCACTCCCCCGTCCCCGTCCCTGCAGCGCCGGACTCCTGNCAGGAGCCCCTGGCAGACC  673

Qy    161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg  180
           |||||||:::||||||||||||||||||||
Db    674  CTCTCCCCACACGGCACAGCTTGCCTAGCACCTTCACCAGCAGCCCCCGTGGCTCTGAGG  733

Qy    181  SerThrIleLeuSerMetSerArgThrThrCysArg  193
           |||||:::||||||||:::||
Db    734  AGTACTATTCCTTCCATGAATCGGACTTGGACTGCCCGA  772
```

RESULT 15
BI411303
LOCUS   BI411303   1040 bp   mRNA   linear   EST 14-AUG-2001
DEFINITION   602964692F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5119942 5',
             mRNA sequence.
ACCESSION   BI411303
VERSION   BI411303.1 GI:15172226
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1040)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
             Bonaldo, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by:Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11292 row: h column: 23
High quality sequence start: 29
High quality sequence stop: 909.

FEATURES   Location/Qualifiers
     source   1..1040
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Czech II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5119942"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTCTGTTTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

```
Alignment Scores:
Pred. No.:                              Length:               1040
Score:                   3.27e-24       Matches:              580.00
Percent Similarity:      67.3%          Conservative:         130
Best Local Similarity:   62.5%          Mismatches:           10
Query Match:             51.1%          Indels:               9
DB:                                     Gaps:                 3
                                                              2

US-09-989-890-238 (1-212) x BI411303 (1-1040)

Qy      1  SerProHisGlnAlaAlaAlaAlaProValAspProArgThrLeuAlaThrMetGly  20
           ||||||||||::|||
Db    346  AGCCCCCACCAGGGCGGTCGCACCCGGAGGCCAGACCCTAAGGACCCTGGCCACCACGGGGC  405

Qy     21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla  40
           |||||||||||||||::::||||||||
Db    406  CAGAGAGTATCACCTTCATTTCAGGCTCTGCAGAACCAACAGCCCCAACCTGCT  465

Qy     41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuGlyLeuProSerAlaSerAlaAla  60
           ||||||||||::|||:::||||||||||||||
Db    466  GCCTCCTCTGGGCGCCCCCTGGGGTTGGACTGGTGTAGGGCTGCCTTCTTGCTTCCGACGCT  525

Qy     61  AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80
           ||||||||||||||||||:::||||||||||||||||
Db    526  GCAGGGAATTGCCTCCTGCCGCTGGAGCTTGTGTGCGGGGCTGCAGCCCCTGCTTATCTG  585

Qy     81  LeuArgThrProLeuArgGlyLeuLeuSerProThrGlyProArgSerThrMetGluCys  100
           |||||||||:::|||||||::::|||||||||||||
Db    586  CCGGAGACCCCATTGAAGGGTCTGCCGAAGCCGCTGGGCCCAAGGAACAATGTGTGC  645

Qy    101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPro  120
           |||||||||||||||||||||||||::::||
Db    646  CCCCCAGCCCCGGACCGTGCACCCCCAGCCGCGCGGGATGGCCGGAGGCTCAAGACCAGCA  705

Qy    121  TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
           ||||||||||:::||||||||||:::|||
Db    706  TGGGCAGCAGCTTCAGCTACCCTGATGTTAAGCTCAAAGGCATCCCTGTCTACCCCTACC  765

Qy    141  ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
           |||||||||||:::|||:::|||
Db    766  GCCATGCCACTTCCCAGTCCCTGACTGTCCGCTGCAGGAGCCCTGCTGCAAGGAGCCCGGAGC  825

Qy    161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg  180
           |||                :::|||
Db   ...
```

```
Db    826 CTCCTCCCACACGGCACAGCTTGCCTAGCAACTTCACCA--------AACAGCCCCCCG 876

Qy    181 SerThrIle---------LeuSerMetSerArgThrArgThrTrpThrCys-ArgArgTr 195
              :::        :::||||||||||||||     |||||
Db    877 CGGCTCCTTGAGGAGTACTACTCCCTTCCATGAACTCGGACCTGCCCTGAGATG 936

Qy    195 pAlaValAlaProCysArgAla 202
              |||| ||||||||||||||
Db    937 GGCAGGGGCTCCATGTCGAGCC 958
```

Search completed: March 17, 2006, 07:14:01
Job time : 3879 secs

THIS PAGE LEFT BLANK

```
                    GenCore version 5.1.7
          Copyright (c) 1993 - 2006   Biocceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      March 17, 2006, 05:36:01 ; Search time 350 Seconds
                                        (without alignments)
                                        1076.694 Million cell updates/sec

Title:       US-09-989-890-238
Perfect score: 1134
Sequence:    1 SPHQAAAPVDQTPRTLATMG..........RRWAVAPCRAEKLMCSSSRS 212

Scoring table:  BLOSUM62
             Xgapop  10.0 , Xgapext   0.5
             Ygapop  10.0 , Ygapext   0.5
             Fgapop   6.0 , Fgapext   7.0
             Delop    6.0 , Delext    7.0
             Listing first 45 summaries

Searched:    1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters:        2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095249_16649/app_query.fasta_1
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs808 -USER=US09989890 @CGN 1_1 193 @runat 16032006_095249 16649
-NCPU=6 -ICPU=3 -NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Issued_Patents_NA:*
             1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
             2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
             3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
             4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
             5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
             6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
             7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
             8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
             9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Match Length	DB	ID	Description
1		1128	99.5	1785	3	US-10-104-047-799	Sequence 799, App
2		178	15.7	1716	3	US-09-902-540-9034	Sequence 9034, Ap
3		178	15.7	9198	3	US-09-902-540-966	Sequence 966, App
4		168.5	14.9	2703	3	US-09-902-540-2939	Sequence 2939, Ap
5		168.5	14.9	23417	3	US-09-902-540-1207	Sequence 1207, Ap
c 6		166	14.6	71989	3	US-09-443-501A-2	Sequence 2, Appli
c 7		166	14.6	71989	3	US-09-724-889A-2	Sequence 2, Appli
c 8		166	14.6	71989	3	US-09-724-876-2	Sequence 2, Appli
c 9		165	14.6	68750	3	US-09-335-409-1	Sequence 1, Appli

c 10		165	14.6	68750	3	US-09-568-102-1	Sequence 1, Appli
c 11		165	14.6	68750	3	US-09-567-969-1	Sequence 1, Appli
c 12		165	14.6	68750	3	US-09-568-480-1	Sequence 1, Appli
c 13		165	14.6	68750	3	US-09-568-486-1	Sequence 1, Appli
c 14		165	14.6	68750	3	US-09-568-472-1	Sequence 1, Appli
c 15		165	14.6	68750	3	US-09-567-899-1	Sequence 1, Appli
c 16		165	14.6	68750	3	US-10-014-717-1	Sequence 1, Appli
c 17		164	14.5	13438	3	US-09-949-016-13231	Sequence 13231, A
c 18		159.5	14.1	4079	3	US-09-016-434-1248	Sequence 1248, Ap
c 19		159.5	14.1	4517	3	US-09-949-016-4573	Sequence 4573, Ap
c 20		159.5	14.1	4519	3	US-09-016-434-1202	Sequence 1202, Ap
c 21		159.5	14.1	4519	3	US-09-949-016-650	Sequence 650, App
c 22		159.5	14.1	4519	3	US-09-492-027-3	Sequence 3, Appli
c 23		159.5	14.1	6935	3	US-09-902-540-865	Sequence 865, App
c 24		158.5	14.0	25048	3	US-09-902-540-1239	Sequence 1239, Ap
c 25		158	13.9	1275	3	US-09-902-540-3545	Sequence 3545, Ap
c 26		158	13.9	17228	3	US-09-902-540-1170	Sequence 1170, Ap
c 27		157.5	13.9	2699	3	US-09-902-540-6501	Sequence 6501, Ap
c 28		157.5	13.9	2700	3	US-09-902-540-467	Sequence 467, App
c 29		156.5	13.8	1092	3	US-10-439-479-9	Sequence 9, Appli
c 30		156.5	13.8	2465	3	US-10-439-479-24	Sequence 24, Appl
c 31		156	13.8	1191	3	US-09-902-540-3329	Sequence 3329, Ap
c 32		156	13.8	19302	3	US-09-902-540-1155	Sequence 1155, Ap
c 33		156	13.8	28958	2	US-08-258-261B-6	Sequence 6, Appli
c 34		156	13.8	28958	2	US-08-456-837-6	Sequence 6, Appli
c 35		156	13.8	28958	2	US-08-457-342-6	Sequence 6, Appli
c 36		156	13.8	28958	2	US-08-457-646A-6	Sequence 6, Appli
c 37		156	13.8	28958	2	US-08-458-076A-6	Sequence 6, Appli
c 38		156	13.8	28958	2	US-08-764-233A-4	Sequence 4, Appli
c 39		156	13.8	28958	2	US-08-457-335A-6	Sequence 6, Appli
c 40		156	13.8	28958	2	US-08-729-214-6	Sequence 6, Appli
c 41		156	13.8	28958	2	US-09-028-934-6	Sequence 6, Appli
c 42		156	13.8	33529	3	US-09-144-085-3	Sequence 3, Appli
c 43		156	13.8	49377	2	US-08-764-233A-1	Sequence 1, Appli
c 44		155	13.7	58857	3	US-09-477-962-1	Sequence 1, Appli
c 45		154	13.6	1432	3	US-09-620-312D-420	Sequence 420, App

ALIGNMENTS

```
RESULT 1
US-10-104-047-799
; Sequence 799, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. 6943241e1 full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE: 2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 799
;  LENGTH: 1785
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-104-047-799

Alignment Scores:
Pred. No.:                           Length:           1785
Score:                 1128.00       Matches:           211
Percent Similarity:    99.5%         Conservative:        0
Best Local Similarity: 99.5%         Mismatches:          1
Query Match:           99.5%         Indels:              0
DB:                     3            Gaps:                0

US-09-989-890-238 (1-212) x US-10-104-047-799 (1-1785)

Qy      1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20

Db    193 AGCCCCCACCAAGCCGGCGGCACCCGGTAGACCAAGACCCCAAGGACCCTGGCCACCATGGGC 252
```

Qy   2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln   21
           ||| |||
Db 1471 CCCCATCCCGCGGCGGCA--------------------CGCCCACCAGCAACCGCTGGTTGT 1430

Qy   22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAlaAla   41
Db 1429 CGGCCGCCACCCGTGCGTCAACACCAGCGTCTCGGCGGCGAAGCCCCGCGCCCTGGGCC 1370

Qy   42 -----CysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla   59
Db 1369 AGGTGGCCAGCCGGAGCTCCT---TGCTCGTCAGGTGCTGGCCTTCCTTCAGCGAGTACA 1313

Qy   60 -----AlaGlyIleAlaSerSerAlaValGluProValCys   72
Db 1312 GCCCGGGCCGGCTTCCAGGTCCGGGAGCCCCTTCAGTCGGTGCGGTGCGGCGGCGTGA 1253

Qy   73 GlyAspAla-----AlaProLeuLeuArg---ThrProLeuArg   86
Db 1252 AACTCAGGCTGTCCACATCCGGTTCACCGTGCGTCCCGACGCGCTCGGCCACCACCACGT 1193

Qy   87 GlyLeuLeuLysProThrGlyProArgSerThrMetGluCysProProAlaLeuIle--- 105
Db 1192 CCCGGGTGGCCACCGCGGCGGCCCAGTTGCTCGGGGAGGTGCTCCCCACCGGTGTTCCGT 1133

Qy 105   105
Db 1132 CCCGGGTGCTCAACGCCTGAAGCCCATTCGGGCGACCAGCAGCCCGTTCACCACCCGCGA 1073

Qy 106 ValHisProAlaGlyGlyMetAlaSerGlySerGlnProTrpAlaAlaAlaSer 125
Db 1072 GCTCGCCCGGCGCAGGGGCTCCGTCCTCCGCAGCCGGCACCC----------- 1028

Qy 126 AlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr--------- 140
Db 1027 ---CGTCCGAGGAGGAGCCATCAGCAACAGAGGGCGTCCCGACGCGGCGGCGCCCG 971

Qy 141   ArgGlyProProGlnPro--- 147
Db 970 TGTTCAGCGTGGGGCCGGAAGGCAACGTAGAGTCCCCACCGGTCCGTCGAGGCGGAGGCAT 911

Qy 148   -----LeuMetArgThrPro 152
Db 910 ACGGGTGCGGGTGGTTGCACTCGGCGAGCATCGGTCCGTCAACTCCGCGCGGCGACACCA 851

Qy 153 AlaAlaArgSerHisTrpProIlePro-----HisProCysAspThrAlaCys----- 168
Db 850 TGTGGCCGAAGCGAT---CCAGCACCACCAGGAAGTACATGCGGCGCGTGTCGCGCGX 794

Qy 169   -----ProAlaProLeuProVal 174
Db 793 GCTCGTGCCGTGGGGTAGGCCTCGAAGAGGGGCCGCCAGCCGGTCCGGGCCATCACCGCCA 734

Qy 175 ValLeuValAlaProArgSerMetSerArgThrIleLeuSerMetSerArgThrCysArgArg 194
Db 733 ---------GCCGCGCCAGAAGACGTCTGCGTCCGTGCCGGAGTCCGCGCCGCGTCGG 680

Qy 195   --------TrpAlaValAlaProCysArgAla 202
Db 679 CCAGGGTGAAGGTCCACCGCGTGTCCCCGTGTCGCGCT 641

Qy   21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAla   40
Db 253 CAGAGAGCATTACCTTCATCTCTCGGCCTCGCTGCTGGAGCCGGCCCTTGAGTCCCCACCTGCT 312

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla   60
Db 313 GCCTGCTCTGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCCGCT 372

Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCysGlySerGlyAspAlaAlaCysLeu   80
Db 373 GCCGGGATTGCCTCCAGCGGCGTCGGAGCCGTGGGGATGCAGCGCGCCCCTGCCTGTCTA 432

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
Db 433 CTGAGGACTCCACTGAGGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC 492

Qy 101 ProProAlaLeuIleValHisProAlaGlyGlyMetAlaSerGlySerGlnPro 120
Db 493 CCCCCAGCCCCTGATCGTGACACCCCCCAGCCCGGCGGGATGGCCAGCAGCTCAAGTCAACCA 552

Qy 121 TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
Db 553 TGGGCAGCAGCTTCAGCTACACCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC 612

Qy 141 ArgGlyProProGlnProProLeuMetArgThrProAlaAlaArgSerHisTrpProIle 160
Db 613 CGAGGGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGTCGCGAAGGAGCCCACTGGCCGATC 672

Qy 161 ProHisProCysAspThrAlaCysAlaAlaProLeuProValValLeuValAlaProArg 180
Db 673 CCCCACCCATGCGACACAGCCTGCCCAGCACCCTTGCCAGTAGTCCTCGTGGCTCCGAGG 732

Qy 181 SerThrIleLeuSerMetSerArgThrCysThrTrpThrCysArgArgTrpAlaValAlaProCys 200
Db 733 AGTACTATTCTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGCAGTTGCTCCATGT 792

Qy 201 ArgAlaGluLysLeuMetCysSerSerSerArgSer 212
Db 793 CGAGCCGAGAAATTGATGTGTCATCTTCAAGGAGC 828

Alignment Scores:

Pred. No.:	0.242	Length:	1716
Score:	178.00	Matches:	84
Percent Similarity:	33.1%	Conservative:	13
Best Local Similarity:	28.7%	Mismatches:	88
Query Match:	15.7%	Indels:	108
DB:	3	Gaps:	15

US-09-989-890-238 (1-212) x US-09-902-540-9034 (1-1716)

; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 966
; LENGTH: 9198
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-966

Alignment Scores:
Pred. No.:   1.54   Length: 9198
Score: 178.00   Matches: 84
Percent Similarity: 33.1%   Conservative: 13
Best Local Similarity: 28.7%   Mismatches: 88
Query Match: 15.7%   Indels: 108
DB: Gaps: 15

US-09-989-890-238 (1-212) x US-09-902-540-966 (1-9198)

Qy   2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
            |||||||| || ||| ||| ||||
Db   1471 CCCATCCGCCGCGGCGGCA----------------------CGCCACCAGCACCGCTGGTTGT 1430

Qy   22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAlaAla 41
            ||| ::: |||
Db   1429 CGGCGCCACCGCGTGCGTCAACACCAGCGTCTCGGGGGAAGCCCCGCGCCCCTGGGCC 1370

Qy   42 ------CysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla 59
                :::|| || ||
Db   1369 AGGTGGCCAGCGGCGGAGCTCCT---TGCTCGTCAGGTGCTGCGCCTTCCTTCAGCGAGTACA 1313

Qy   60 Ala-------------------AlaGlyIleAlaSerSerAlaValGluProValCys 72
            ||| ||||| |||
Db   1312 GCCACGCCGGCTCCAGGTCCGGGAGCCGGAAGCCCTTCAGCTGCGTGGCGCCGGCGTGA 1253

Qy   73 GlyAspAla------------------AlaProAlaCysAlaProLeuLeuArg---ThrProLeuArg 86
            |||||| ::|||| |||
Db   1252 AACTCAGGCTGTCCACATCCGCGTTCACCGTGGTGCCGACGCGCTGGGCACCCACCGT 1193

Qy   87 GlyLeuLeuLysProThrGlyProArgSerThrMetGluCysProProAlaLeuIle--- 105
            |||||||| ||| :::
Db   1192 CCCGGGTGGCCACCGGCGGCGCAGTTGCTGGGCGAGGTGCTCCCCCACCGGTGTTCCGT 1133

Qy   105 --- 105
Db   1132 CCCGGGTGCTCAACGCCTGGAAGCCCCATTCGGCAGCAGCAGCCCGTTCACCACCGCGA 1073

Qy   106 ValHisProProAlaGlyIleGlyMetAlaSerGlySerSerGlnProTrpAlaAlaAlaSer 125
            ||| |||||||| ::||||| |||||
Db   1072 GCTCCCGCCCGGCGCAGGGGCTCCTCCGCAGCGGCGGCACCC------------------ 1028

Qy   126 AlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr----------------- 140
            ||||::: |||
Db   1027 ---CGTCCGAGGAGAAGGCCCATCAGCAACGAGGGGCGTCCCCGACGCGGCGCGCCCG 971

Qy   141 ---------------------ArgGlyProProGlnPro--------------------- 147
                                 ||||||:::
Db   970 TGTTCAGCGTGGGCGCGAAGGCAACGTAGAGGTCCCCACCCGGCGTCCGAGGGCGAGGCCAT 911

Qy   148 ----------------------------------LeuMetArgThrPro 152
                                              ||||||||
Db   910 ACGGGTGGCGGTGGTTGCACTCGGCGCAGCATCGGCGGCCGTCAACTCGGCGGGCGGACACCA 851

Qy   153 AlaAlaArgSerHisTrpProIlePro------HisProCysAspThrAlaCys------- 168
            ||| ::: |||
Db   850 TGTGGCGAAGGCAT---CCAGCACCACCAGGAGTACATGCGGACGTACATGGCGGCCGTGTCGCGCC 794

Qy   169 -----------------------ProAlaProLeuProVal 174
                                   |||||| :::
Db   793 GCTGGTCCGTGGGGTAGGCCTCGGAGAGGGGCCGCCAGCCGGTCCGGCGCCATCACCCGCCA 734

Qy   175 ValLeuValAlaProArgSerThrIleLeuSerMetSerArgThrTrpCysThrCysArgArg 194
            |||||| :::
Db   GCTCTGCGCC------GGTGTCCGCCA

Qy   64 AlaSerSerAlaAlaValGluProValCysGly------------------------
                ::: ||| |||
Db 12818 ACAGGCCTCCGGTGCACCTCGCGGCAGGCGTGCCGTTGGCGACCGGTCGGTCGTGTGCTGCT 12877

Qy   76 AlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArg   95
              ||| ||| |||
Db 12878 GCGCCCAACTGCTGGCATAGGCATCCAACACCTGCC----------CACCCGCGCCCG--- 12925

Qy   96 SerThrMetGluCysProAlaLeuIleValHisProProAlaGlyGlyMetAlaSer   115
                 |||
Db 12926 ----------CCATGTCCC-----------------------CCGGCAGGGGGC   12946

Qy   116 GlySerSerGlnProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSer   135
              ||| ||| :::
Db 12947 ---TGCCCGTGGCCGCGAAGGCGGAGGACAGTCGCTGCTTCGCCCTCGGCCCCCCC   12997

Qy   136 LeuCysIleProThrArgGlyProProGlnProLeuMetArg---------------   150
                |||
Db 12998 AGGTGTCCGCCA------GCAGCCCCTCCGGCCGCCAGACGCTGGACTGAAACCACA   13051

Qy   151 ----------ThrProAlaAlaArgSerHisTrpProIlePro------HisProCys   164
                                                          ||||||
Db 13052 GCCGGCCTCCCACCGCCCCCACCAGCGCCACCGTGCCGCACCGGCCATGCTCCCGCCTTGC 13111

Qy   165 Asp------------------------------ThrAlaCysProAla------   170
            ||| ||||||
Db 13112 GGGCCCGTGCGCGTGCTCCTTGCGACAGCGCCTCCAGCAGCGCCCCCATGGAGGCGA   13171

Qy   171 ------------ProLeuProValValLeu------------ValAlaProArgSerThr   182
                        |||||||||
Db 13172 AGCGCGCCTCCGGCTCCAGTGACAACCCCCGCATCACCGCGCCCGCACCCAGGCGG---   13228

Qy   183 IleLeuSerMetSerArgThrThrCysArgArgTrpAlaValAlaProCysArgAla   202
             |||
Db 13229 ------GGACGCGGCGGCGTCGCTGGGCGGCTCGTGGATGGGCGGCGCGAGGGCTCCCGCCCCGGCTC 13282

Qy   203 GluLysLeuMetCysSerSerSerArg   211
             |||
Db 13283 GCTCCAGCGCGGCGGCCGTCACCTTCGCGG   13309

RESULT 6
US-09-443-501A-2/c
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Qy   151 ----------ThrProAlaAlaArgSerHisTrpProIlePro------HisProCys   164
                                                          ||||||
Db 1075 GCCCGGCCTCCCACCGCCCCCACCAGCGCCACCGTGCCGCGGCCAATGCTCCCGCCTTGC 1016

Qy   165 Asp------------------------------ThrAlaCysProAla------   170
            ||| ||||||
Db   956 GGGCCCCGTGCGCGTGCTCCTTGCGACAGCGCCTCCAGCAGCGCCCCCATGGAGGCGA 1015

Qy   171 ------------ProLeuProValValLeu------------ValAlaProArgSerThr   182
                        |||||||||
Db   899 AGCGCGCCTCCGGCTCCAGTGACAACCCCCGCATCACCGCGCCCGCACCCAGGCGG--- 955

Qy   183 IleLeuSerMetSerArgThrThrCysArgArgTrpAlaValAlaProCysArgAla   202
             |||
Db   845 ------GGACGCGGCGGCGTCGCTGGGCGGCTCGTGGATGGGCGGCGCGAGGGCTCCCGCCCCGGCTC 898

Qy   203 GluLysLeuMetCysSerSerSerArg   211
             |||
Db   818 GCTCCAGCGCGGCGGCCGTCACCTTCGCGG   844

RESULT 5
US-09-902-540-1207
; Sequence 1207, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1207
; LENGTH: 23417
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(23417)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1207

Alignment Scores:
Pred. No.:   Length: 23417
Score: 14.1   Matches: 79
                     168.50   Conservative: 5
Percent Similarity: 31.2%   Mismatches: 100
Best Local Similarity: 29.4%   Indels: 85
Query Match: 14.9%   Gaps: 15
DB: 3

US-09-989-890-238 (1-212) x US-09-902-540-1207 (1-23417)

Qy   1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly   20
             ||||||| ||| |||
Db 12584 TCACCTGCGACAGCGGCGCGCTCCCAATGACTCCAGCTCCGCGGCATTGCCG------GGT 12637

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerPro----------   37
              :::
Db 12638 CCGCGGGGCCCGGGGCTGCTGGCTCCGCCGGGGACTCCACGTCCGCGCAGTCCTCGGGTGAAG 12697

Qy   38 ----------ProProAlaAlaCysSerGlyAspProGlyCysGlySerGlyAlaGly   53
                        |||
Db 12698 GCAGCGCATACGCCCGCGCCGCCTTCTCCACGCCCGCCTTGTCCACCTGGGACAGGG 12757

Qy   54 LeuProSerAlaSerAla-----------------------AlaAlaGlyIleIle   63
             ||||||| |||
Db 12758 CGCCCACCACCAGCGCACGCAAATCCTTGTGCCCGCCGCCTCCAGACACCACCGCGCCTGGGTGA 12817

```
; APPLICANT: Ziemann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 3062-20031.00
; CURRENT APPLICATION NUMBER: US/09/724,889A
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/443,501A
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-724-889A-2
```

```
Alignment Scores:
Pred. No.:                                  Length:          71989
Score:                  66.5                Matches:            72
Percent Similarity:     166.00              Conservative:       17
Best Local Similarity:  36.5%               Mismatches:        103
Query Match:            29.5%               Indels:             52
DB:                     14.6%               Gaps:                9
```

```
US-09-989-890-238 (1-212) x US-09-724-889A-2 (1-71989)

Qy      2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
            |||    ||||    ||    :::||||||
Db  18069 CCCACGACCGCCACAGCGCCGTACTCCACCGGCGTACTCCACCGGCGGGGGCTGGG--- 18013

Qy     22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAlaAla 41
            ||||||              ||||||          :::
Db  18012 ---TGAACGCCGTCTGGTCCAGCAACAACGACTCGGCGCCTCCCGCCCACATCA 17956

Qy     42 -----CysSerGlyAspProGlyCysGlySerGlyAlaGly- 53
                  :::||||      ||
Db  17955 CCTCGGCCAGCGGCGGCGGTCCCGGTCGAACAGCGTCACGCACCGGTCGAACGCCT 17896

Qy     54 -------LeuProSerAlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluPro 70
                   |||          ||          :::||
Db  17895 CCCGGAACGCTGGCCACGCCGCAGAGCCCCGGCGCCGGTCTGCGCGCCCT 17836

Qy     71 ValCysGlyAspAlaAlaProAlaCysLeuLeuArgThrProLeuLeuLeuLys 90
            ||||      ||      |||:||||
Db  17835 GTCCGGTGAACAGCAACGCCAGCTTGCCGCGCGAGGAGC--------- 17797

Qy     91 ProThrGlyProArgSerThrMetGluCysProProAlaLeuIleValHisProProAla 110
            ||||:||||||          ||||||
Db  17796 ----------------TCGGCGATGCAGCGCGCCGCCC--------------CCGCCGGCG 17767

Qy    111 GlyGlyMetAlaSerGlySerSerGlnProTrpAlaAlaAlaSerAlaThr------Pro 128
            ||||||                          ||||||
Db  17766 TCTGCCCCTGCGCCACGCCGCCGAAAGCCGGCCCAGCAGCCCCTCGCGCGACGTCACCGCCA 17707

Qy    129 MetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProProProGlnProLeu 148
            :::                    :::   :::||
Db  17706 CCGGCGAGCCGGTGGGTCATCGCGGCGCTGCGCGGTCGTCGCCAGCGTGAACGCTCCCCGA 17647

Qy    149 MetArgThr--------------ProAlaAlaArgSerHisTrpPro--------- 159
            |||                    |||     |||:||
Db  17646 GCCCGAGCTCCGGGTGCATGTCCAGGTGCTCGCGCACCGCCGCCTGCCGGTCGGAGCG 17587

Qy    160 IleProHisProCysAspThrAla------------------------------------Cys 168
```

```
Qy    54  --------LeuProSerAlaSerAlaAlaAlaAlaGlyIleAlaSerAlaValGluPro  70
                            ::|||               |||          ::: |||
Db  17895  CCCGGACGGTGGCCACGCGCAGAGCCCCGGCCATGCCGGCGTCTGCGCGCCCT  17836

Qy    71  ValCysGlyAspAlaAlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLys  90
              ||| |||  |||  |||||| :::
Db  17835  GTCCGGTGAACAGCAACGCCAGCTTGCCGCCGAGAGC----------------  17797

Qy    91  ProThrGlyProArgSerThrMetGluCysProAlaLeuIleValHisProProAla  110
              ||| |||||                        ||||||||
Db  17796  --------TCGGATGCAGCGCGC-CGCCGCCGCC----------CCGCCGGCG  17767

Qy   111  GlyGlyMetAlaSerGlySerGlnProTrpAlaAlaSerAlaThr-------Pro  128
              |||                        ||:|||||
Db  17766  TCTGCCCCTGCGCCACGGCCGAAAGCCGCCAGCCCCTGCGCACGACGTCACCGCCCA  17707

Qy   129  MetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProProGlnProLeu  148
              :::  ||| |||            ||:      |||||::::||
Db  17706  CCGCGGAGCCGGTGGGTCATCGGCCGTCGTCGCCAGGTGACGCCACGTCCCGA  17647

Qy   149  MetArgThr-----------ProAlaAlaArgSerHisTrpPro--------  159
              ||| |||             ||  ||||||:::
Db  17646  GCCCGAGCTCCGGGTGCATGTCCAGGTGCTCGGCAGCCGGCCGCCCTGCGCGTCGAGCG  17587

Qy   160  IleProHisProCysAspThrAla----------------Cys  168
              |||  |||||| ||
Db  17586  CCCCCTCGCTCTTGCCCGACAGCACCAAAAGCTCCGCCGAGCGGCTCCGGCGGCCAGGCC  17527

Qy   169  ProAlaProLeuProValValLeuValAlaProArgSerThrIleLeuSerMetSerArg  188
              |||||  :::|||  |||||| :::
Db  17526  ACAGCTCCACCGCCGCGCCTTCCAGCACCACATGCGCGTTCGTTCCGCTCATCCCGA  17467

Qy   189  ThrTrpThrCysArgArgTrpAlaValAlaProCysArgAlaGluGluLysLeuMetCysSer  208
              |||   |::|                      |||
Db  17466  ACGAGCTCACCCCCGCGAAGCGCGGACGGTCCGTGCGCGCGCACCGGCACCGGCTCGGTCG  17407

Qy   209  SerSerArgSer  212
              :::|||:::
Db  17406  CCAACGCGAGCG  17395
```

```
RESULT 8
US-09-724-876-2/c
;  Sequence 2, Application US/09724876
;  Patent No. 6921650
;  GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT: Julien, Bryan
;  APPLICANT: Katz, Leonard
;  APPLICANT: Khosla, Chaitan
;  APPLICANT: Tang, Li
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION: Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/724,876
;  CURRENT FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: US/09/443,501
;  PRIOR FILING DATE: 1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 2
;    LENGTH: 71989
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Synthetic construct
US-09-724-876-2

Alignment Scores:
Pred. No.:                              Length:        71989
Score:              66.5                Matches:          72
Percent Similarity: 166.00              Conservative:     17
Best Local Similarity: 36.5%            Mismatches:      103
Query Match:        29.5%               Indels:           52
DB:                 14.6%               Gaps:              9
                     3
```

```
US-09-989-890-238 (1-212) x US-09-724-876-2 (1-71989)

Qy     2  ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln  21
              :::|||||| |||                    ::: |||
Db  18069  CCCACGACCGCCCACAGCCGCCGTCAGCGGCTACTCCACCGGAAGAGCGCGGCGTGGG---  18013

Qy    22  ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAlaAla  41
              |||||||||                      :::
Db  18012  ---TGAACGCGTCTGGTCCAGCAACAACGACTCGGCGCTCCCCGCCTCCGCCCCACATCA  17956

Qy    42  -----CysSerGlyAspProGlyCysGlyCysGlyAlaGly--------------  53
              |||||||  ||  |||  |||
Db  17955  CCTCGCGCAGCGGCGGCCGGTCGAGCTCCAGCTCCGGTCGAACAGCGTCACCGCACCGGTGACGCGCCT  17896
```

```
Db  17586  CCCCTCGCTCTTGCCCGACAGCACCAAAAGCTCCGCCGAGCGGCTCCGGCGGCCAGGCC  17527

Qy   169  ProAlaProLeuProValValLeuValAlaProArgSerThrIleLeuSerMetSerArg  188
              |||||  :::|||  |||||| :::
Db  17526  ACAGCTCCACCGCCGCGCCTTCCAGCACCACATGCGCGTTCGTTCCGCTCATCCCGA  17467

Qy   189  ThrTrpThrCysArgArgTrpAlaValAlaProCysArgAlaGluGluLysLeuMetCysSer  208
              |||   |::|                      |||
Db  17466  ACGAGCTCACCCCCGCGAAGCGCGGACGGTCCGTGCGCGCGCACCGGCACCGGCTCGGTCG  17407

Qy   209  SerSerArgSer  212
              :::|||:::
Db  17406  CCAACGCGAGCG  17395
```

```
RESULT 9
US-09-335-409-1/c
;  Sequence 1, Application US/09335409
;  Patent No. 6121029
;  GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 1
;    LENGTH: 68750
;    TYPE: DNA
;    ORGANISM: Sorangium cellulosum
US-09-335-409-1

Alignment Scores:
Pred. No.:                              Length:        68750
Score:              71.6                Matches:          70
Percent Similarity: 165.00              Conservative:     17
Best Local Similarity: 35.7%            Mismatches:      105
Query Match:        28.7%               Indels:           52
DB:                 14.6%               Gaps:              8
                     3
```

```
US-09-989-890-238 (1-212) x US-09-335-409-1 (1-68750)
```

```
;       TYPE: DNA
;       ORGANISM: Sorangium cellulosum
US-09-568-102-1

Alignment Scores:
Pred. No.:
Score:                    71.6          Length:        68750
Percent Similarity:      35.7%          Matches:          70
Best Local Similarity:   28.7%          Conservative:     17
Query Match:             14.6%          Mismatches:      105
DB:                                     Indels:           52
                                        Gaps:              8

US-09-989-890-238 (1-212) x US-09-568-102-1 (1-68750)

Qy      2 ProHisGlnAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln   21
            :::|||||                         :::||||||
Db  23681 CCCACGACCGCCACAGCGCCGTCAGCGGTACTCCACCGCGAAGAGACGCGGGCTGGG---  23625

Qy     22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAlaAla    41
            |||||||||                   :::|||
Db  23624 ---TGAACGCCGTCTGGTCGAGCAACAACGACTCGGCGCTCCCCGCCTCCGCCACATCA    23568

Qy     42 -----CysSerGlyAspProGlyCysGlySerGlyAla                       52
                :::|||||                |||
Db  23567 CCTCGGCGCAGCGGCGGCCGGTCGAGCTCCCGGTCGAACAGCGCCACGACCGGTCGAACGCCT  23508

Qy     53 ------GlyLeuProSerAlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluPro  70
                |||||||
Db  23507 CCCGGAACGCTGGCCACGCGCCAAAGCCCCGGCCATGCCCGGCGTCTGCGCGCCCT      23448

Qy     71 ValCysGlyAspAlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLys    90
             ||||||
Db  23447 GTCCGGTGAACAGGAACGACCAGCTTGCGCGGCGAGGAGC--------------------  23409

Qy     91 ProThrGlyProArgSerThrMetGluCysProProAlaLeuIleValHisProProAla 110
            :::|||
Db  23408 -----------------TCGCGGATGCAGCGGCGGCCGCCC------------CCGCGCGGCG 23379

Qy    111 GlyGlyMetAlaSerGlySerSerGlnProTrpAlaAlaAlaSerAlaThr------Pro 128
             ||||||
Db  23378 TCTGCCCCTGGCCACGCGCCGAGAGCCCCGAGAGCGCGCGCAGCCCCTCGGCGACGTCACCGCCA 23319

Qy    129 MetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProProGlnProLeu    148
             :::|||                |||||||
Db  23318 CCGGCGAGCCGGTGGCTCATCGCGCTGCGCGTCGTCGCCAGGCTGAACGCACGTCCCCGA  23259

Qy    149 MetArgThrProAlaAla-------------------------ArgSerHisTrpPro   159
             |||                                      :::|||
Db  23258 GCCCGAGCTCCGGGTGCATGTCCAGTGCTCGCGGCAGCGCCGCCGCCTGCCATCGAGCG   23199

Qy    160 IleProHisProCysAspThrAla-------------------------Cys        168
             |||||                                         |||
Db  23198 CCCCCTCGCTCTTGCCCACAGCACCAAAAGCTCCGCCGAGCGCTCCGGCGGCCCAGGCC   23139

Qy    169 ProAlaProLeuProValLeuValLeuValAlaProArgSerThrIleLeuSerMetSerArg 188
             ||||||                              :::
Db  23138 ACAGCTCCACCGCCCGGCGCCCTCTTCCAGCACCACCATGGCGGTTCGTTCCGCTCATCCCGA 23079

Qy    189 ThrTrpThrCysArgArgTrpAlaValAlaAlaProCysArgAlaGluLysLeuMetCysSer 208
             |||   :::                      |||
Db  23078 ACGAGCTCACCCCCGCGAAGCGCGGGCGGTCCGTGCGCGGCCACGGCACCGGTCGGTCG   23019

Qy    209 SerSerArgSer 212
             :::||||
Db  23018 CCAACGCGAGCG 23007
```

RESULT 10
US-09-568-102-1/c
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750

RESULT 11
US-09-567-969-1/c
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

Alignment Scores:
Pred. No.:
Score: 71.6
Percent Similarity: 165.00
Best Local Similarity: 35.7%
Query Match: 28.7%
DB: 14.6%

Length: 68750
Matches: 70
Conservative: 17
Mismatches: 105
Indels: 52
Gaps: 8

US-09-989-890-238 (1-212) x US-09-567-969-1 (1-68750)

Qy   2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
            ::: |||||||
Db 23681 CCCACGACCGCCACAGCGCCGTCAGCGGTACTCCACCCGGAAGAGCGCGGGCTGGG-- 23625

Qy   22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAlaAla 41
            ::: ||||
Db 23624 ---TGAACGCCGTCTGGTCGAGCAACAACGACTCGGGCCTCCCGCCCACATCA 23568

Qy   42 -----CysSerGlyAspProGlyCysGlySerGlyAla 52
                 |||
Db 23567 CCTCGCGCAGCGGCGGGTCAGCTCCCGGTCGAACAGCGCCACGGTCGAACGCGT 23508

Qy   53 ------GlyLeuProSerAlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluPro 70
            ::: |||||||
Db 23507 CCCGGAACGCTGGCCACGCGCGCAAAGCCCCCGCCCATGCCCCGGCGTCTGCGCCCCT 23448

Qy   71 ValCysGlyAspAlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLys 90
            ::: |||||||||
Db 23447 GTCCGGTGAACAGGAACGGCCAGCTTGCCGCGCGAGGAGC- 23409

Qy   91 ProThrGlyProArgSerThrMetGluCysProAlaLeuIleValHisProProAla 110
            ||||||||
Db 23408 -----------------TCGCGATGCAGCGGCCGGCGCC- 23379

Qy   111 GlyGlyMetAlaSerGlySerSerGlnProTrpAlaAlaAlaSerAlaThr-----Pro 128
            |||||||
Db 23378 TCTGCCCCTGCGCCACGCGGCCGAGAGCGCCGCCAGCAGCCCCTCGCGACGTCACCGCCA 23319

Qy   129 MetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProProGlnProLeu 148
            ::: ||||||||
Db 23318 CCGGCGAGCCGGCGCCGCCGCGCTCGCCGCCGGTCGTCGCGCCGCCAGCGTGAACGCCGA 23259

Qy   149 MetArgThrProAlaAla-------------------------------ArgSerHisTrpPro 159
            ||| :::||||
Db 23258 GCCCAGCTCCGGGTGCATGTCCAGGTGGTCGCAGGTGCTCGCCAGCGCCCTGCCGCATCGAGCG 23199

Qy   160 IleProHisProCysAspThrAla----------------------Cys 168
            |||||| :::
Db 23198 CCCCCTCGCTCTTGCCGACAGCACCAAAAGTCCGCCGGCGCTCCGGCGCGGCAGGCC 23139

Qy   169 ProAlaProLeuProValValLeuValAlaProArgSerThrIleLeuSerMetSerArg 188
            ||| :::
Db 23138 ProAlaProLeuProValValLeuValAlaProArgSerThrIleLeuSerMetSerArg
Db 23138 CCGGCAGCCGGCCGCCGCCGCCTCTTCGCACGCACCACATGCGGCTGTTCCGCTCATCCCGA 23079

Qy   189 ThrTrpThrCysArgArgTrpAlaValAlaAlaProCysArgArgAlaGluAlaLysLysCysSer 208
            ||| ||||||
Db 23078 ACAGCTCACCCCGCCGCCCGCGCGCTCGCCGGTCCTCGCCGGTCGTCCGGTCGTCCGG 23019

Qy   209 SerSerArgSer 212
            :::||||
Db 23018 CCAACCGCAGCG 23007

RESULT 12
US-09-568-480-1/c
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

Alignment Scores:
Pred. No.:   71.6
Score: 165.00
Percent Similarity: 35.7%
Best Local Similarity: 28.7%
Query Match: 14.6%
DB: 3

Length: 68750
Matches: 70
Conservative: 17
Mismatches: 105
Indels: 52
Gaps: 8

US-09-989-890-238 (1-212) x US-09-568-480-1 (1-68750)

Qy   2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
            ::: |||||||
Db 23681 CCCACGACCGCCACAGCGCCGTCAGCGGTACTCCACCCGGAAGAGCGCGGGCTGGG-- 23625

Qy   22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAlaAla 41
            ::: ||||
Db 23624 ---TGAACGCCGTCTGGTCGAGCAACAACGACTCGGGCCTCCCGCCCACATCA 23568

Qy   42 -----CysSerGlyAspProGlyCysGlySerGlyAla 52
                 |||
Db 23567 CCTCGCGCAGCGGCGGGTCAGCTCCCGGTCGAACAGCGCCACGGTCGAACGCGT 23508

Qy   53 ------GlyLeuProSerAlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluPro 70
            ::: |||||||
Db 23507 CCCGGAACGCTGGCCACGCGCGCAAAGCCCCCGGCAAAGCCCCGGCGTCTGCGCCCCT 23448

Qy   71 ValCysGlyAspAlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLys 90
            ::: |||||||||
Db 23447 GTCCGGTGAACAGGAACGGCCAGCTTGCCGCGCGAGGAGC- 23409

Qy   91 ProThrGlyProArgSerThrMetGluCysProAlaLeuIleValHisProProAla 110
            ||||||||
Db 23408 -----------------TCGCGATGCAGCGGCCGGCGCC- 23379

Qy   111 GlyGlyMetAlaSerGlySerGlnProTrpAlaAlaAlaSerAlaThr------CCGCCGGCG 128
            |||
Db 23378 TCGGGATGCAGCGGCGAGCGGCCGCCGCC------------------CCGCCGGCG 23379

Qy   129 MetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProProGlnProLeu 148
            ::: ||||||||
Db 23378 TCTGCCCCTGCGCCACGCGGCCGAGCAGCAGCCCCTCGCGACGTCACCGCCA 23319

Db 23318 CCGGCGAGCCGGCGCCGCTGCTCATCGCCGTCGTCCGGTCGTCGCGCCGCCAGCGTGAACGCCGA 23259

```
Db 23447 GTCCGGTGAACAGGAACGCCAGCTTGCCGCGGAGGAGC---              23409
            |||                  |||::::
Qy    91 ProThrGlyProArgSerThrMetGluCysProAlaLeuIleValHisProProAla  110
Db 23408 --------------TCGCGATGCAGCGGCCGCCCC--------------CCGCCGGCG 23379
Qy   111 GlyGlyMetAlaSerGlySerSerGlnProTrpAlaAlaSerAlaThr-------Pro 128
Db 23378 TCTGCCCCTGCGCCACGGCCGAGAGGCCGCCGCAGCAGCCCCTCGCGGCGAGCGTCACCGCCA 23319
Qy   129 MetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProProGlnProLeu 148
Db 23318 CCGGGAGCCGCGGTGGCTCATCGCCGCTGCGCGTCGTCGGCCAGGCTGAACGGCCACGTCCCCGA 23259
Qy   149 MetArgThrProAlaAla--------------------ArgSerHisTrpPro 159
Db 23258 GCCCGAGCTCCGGGTGCATGTCCAGTGCTCGCGCAGCCGCCGCCGCCGCCGCATCGAGCG 23199
Qy   160 IleProHisProCysAspThrAla--------------------------Cys 168
Db 23198 CCCCCTCGCTCTTGCCCGACAGCACCAAAAGCTCCGCGAGCGCTCCGGCGGCGCAGGCC 23139
Qy   169 ProAlaProLeuProValValLeuValAlaProArgSerThrIleLeuSerMetSerArg 188
Db 23138 ACAGCTCCACCGCCGGCGCCTCTTCCAGCACCACATGCGCGTTCGTTCGCTCATCCCGA 23079
Qy   189 ThrTrpThrCysArgArgTrpAlaValAlaProCysArgValAlaGluLysLeuMetCysSer 208
Db 23078 ACGAGCTCACCCCCGCGAAGCGCGGCGGGTCCGTGCGCGCGCACCGGCACCGGCTCGGTCG 23019
Qy   209 SerSerArgSer 212
Db 23018 CCAACGGCGAGCG 23007
```

RESULT 14
US-09-568-472-1/c
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Alignment Scores:

Pred. No.:	71.6	Length:	68750
Score:	165.00	Matches:	70
Percent Similarity:	35.7%	Conservative:	17
Best Local Similarity:	28.7%	Mismatches:	105
Query Match:	14.6%	Indels:	52
DB:	3	Gaps:	8

US-09-989-890-238 (1-212) x US-09-568-472-1 (1-68750)

```
Qy    2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
           |||                  :::||:||
Db 23681 CCCACGGACCGCCACAGCCCGTCAGCGCGTACTCCACCGGCGAAGAGCGCGGGGCTGGG--- 23625
```

```
Qy   149 MetArgThrProAlaAla--------------------ArgSerHisTrpPro 159
           |||                  |||::::
Db 23258 GCCCGAGCTCCGGGTGCATGTCCAGTGCTCGCGCAGCCGCCGCCGCCGCCGCATCGAGCG 23199
Qy   160 IleProHisProCysAspThrAla--------------------------Cys 168
           |||      |||
Db 23198 CCCCCTCGCTCTTGCCCGACAGCACCAAAAGCTCCGCGAGCGCTCCGGCGGCGCAGGCC 23139
Qy   169 ProAlaProLeuProValValLeuValAlaProArgSerThrIleLeuSerMetSerArg 188
           |||                  :::
Db 23138 ACAGCTCCACCGCCGGCGCCTCTTCCAGCACCACATGCGCGTTCGTTCGCTCATCCCGA 23079
Qy   189 ThrTrpThrCysArgArgTrpAlaValAlaProCysArgValAlaGluLysLeuMetCysSer 208
           |||                  :::
Db 23078 ACGAGCTCACCCCCGCGAAGCGCGGCGGGTCCGTGCGCGCGCACCGGCACCGGCTCGGTCG 23019
Qy   209 SerSerArgSer 212
           :::||||:::
Db 23018 CCAACGGCGAGCG 23007
```

RESULT 13
US-09-568-486-1/c
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Alignment Scores:

Pred. No.:	71.6	Length:	68750
Score:	165.00	Matches:	70
Percent Similarity:	35.7%	Conservative:	17
Best Local Similarity:	28.7%	Mismatches:	105
Query Match:	14.6%	Indels:	52
DB:	3	Gaps:	8

US-09-989-890-238 (1-212) x US-09-568-486-1 (1-68750)

```
Qy    2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
           |||                  :::||:||
Db 23681 CCCACGGACCGCCACAGCCCGTCAGCGCGTACTCCACCGGCGAAGAGCGCGGGGCTGGG--- 23625
Qy   22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAlaProAlaAla 41
           |||                  :::||
Db 23624 ---TGAACGCCGTCTGGTCGAGCAACAACGACTCCGCGCTCCGCGCTCCCCCCCATCA 23568
Qy   42 -----CysSerGlyAspProGlyCysGlySerGlyAla------------------ 52
           :::
Db 23567 CCTCCGCGCAGCGGCGGGGGTCCAGCTCCCGGTCGAACAGCGCCACCGGTCGAACGCCT 23508
Qy   53 -----GlyLeuProSerAlaSerAlaAlaLaGlyIleAlaSerSerAlaValGluPro 70
           |||     :::
Db 23507 CCCGGAACGCTGGCCACGCCCGCAAAGCCCCCATGCCGCGTCTGCGCGCCCT 23448
Qy   71 ValCysGlyAspAlaAlaAlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLys 90
```

Alignment Scores:
Pred. No.:
Score: 71.6
Percent Similarity: 35.7%
Best Local Similarity: 28.7%
Query Match: 14.6%
DB:

Length: 68750
Matches: 70
Conservative: 17
Mismatches: 105
Indels: 52
Gaps: 8

US-09-989-890-238 (1-212) x US-09-567-899-1 (1-68750)

```
Qy     2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln  21
          ||| ::: ||| |||||| ||   :::||||||||
Db 23681 CCCACGACCGCCACAGCGCGTCAGCGGTACTCCACCGCGGAAGAGCGCGGGCTGGG--  23625

Qy    22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAlaAla  41
          ||||||||                     |||||||||
Db 23624 --TGAACGCCGTCTGGTCGAGCAACAACGACTCGGCGCTCCCCGGCCTCCGCCCACATCA  23568

Qy    42 ------CysSerGlyAspProGlyCysGlySerGlyAla  52
                ::: ||          ||| |||
Db 23567 CCTCGCGCAGCGGGGGGTCCAGCTCCGGTCGAACAGCGCCACGCCGGTCGAACGCCT  23508

Qy    53 ------GlyLeuProSerAlaSerAlaAlaAlaGlyIleAlaSerAlaValGluPro  70
                :::            ||||
Db 23507 CCCGAACGCTGGCCACGCGCCCATGCCCCGGCGCAAAGCCCCGGCGTCTGCGCGCCCT  23448

Qy    71 ValCysGlyAspAlaAlaProAlaCysLeuLeuArgGlyLeuLeuLys  90
          |||||||||| :::                ||||
Db 23447 GTCCGGTGAACAGGAACGGCCCAGCTTGCGCGCGAGAGC-  23409

Qy    91 ProThrGlyProArgSerThrMetGluCysProAlaLeuIleValHisProProAla  110
                                         ||||||
Db 23408 ------------------TCGCGATGCAGCGCGCCGCCC-  23379

Qy   111 GlyGlyMetAlaSerGlySerGlnProTrpAlaAlaAlaSerAlaThr------Pro  128
          ||||||                              ||
Db 23378 TCTGCCCCTGCGCCACGGCCAGCGCCCCAGAGCGCCGGCGTCACCGCCA  23319

Qy   129 MetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProProGlnProLeu  148
          ||| |||                   :::|||
Db 23318 CCGCGAGCTCCGGTGGCTCATCGCGCTGTCGCGTCGTGACGCTGAACGCTCCCCGA  23259

Qy   149 MetArgThrProAlaAla----------------------ArgSerHisTrpPro  159
          ||| |||                                  :::|||
Db 23258 GCCCGAGCTCCGGTGCATGTCCAGGTGCTCGCGCAGCCGCCGCCGCCGCCATCGAGCG  23199

Qy   160 IleProHisProCysAspThrAla----------------------Cys  168
          ||| ||| :::
Db 23198 CCCCCTCGCTCTTGCCCGACAGCACCAAAAGCTCCGCGAGCGCTCCGGCGCGGCAGGCC  23139

Qy   169 ProAlaProLeuProValValLeuValAlaProArgSerThrIleLeuSerMetSerArg  188
          |||||||| |||||        ::: |||
Db 23138 ACAGCTCCACCGCCGGCCGCCTCTTCCAGCACCACATGCGCTCCGCTCATCCCGA  23079

Qy   189 ThrTrpThrCysArgArgTrpAlaValAlaProCysArgAlaGluLysLeuMetCysSer  208
          :::                   |||          ||| |||
Db 23078 ACGAGCTCACCCCCGCGAAGCGCGGGCGGTCCGTGCGCGGCCACCGGCTCGGTCG  23019

Qy   209 SerSerArgSer  212
          :::|| ||:::|
Db 23018 CCAACGCGAGCG  23007
```

RESULT 15
US-09-567-899-1/c
; Sequence 1, Application US/09567899
; Patent No. 6383787
; Prior Application Number: 09/335,409
; Prior Application Filing Date: 1999-06-17
; Number of SEQ ID NOS: 30
; Software: PatentIn Ver. 2.0

SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Search completed: March 17, 2006, 10:44:58
Job time : 410 secs

```
                    GenCore version 5.1.7
          Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:        March 17, 2006, 05:41:29 ; Search time 780 Seconds
                                   (without alignments)
                                   2247.575 Million cell updates/sec

Title:          US-09-989-890-238
Perfect score:  1134
Sequence:       1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table:  BLOSUM62
                Xgapop  10.0 , Xgapext   0.5
                Ygapop  10.0 , Ygapext   0.5
                Fgapop   6.0 , Fgapext   7.0
                Delop    6.0 , Delext    7.0

Searched:       9793542 seqs, 4113689005 residues

Total number of hits satisfying chosen parameters:    19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095252_16710/app_query.fasta_1
-DB=Published_Applications_NA_Main -QFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss05h
-USER=US09989890 @CGN_1_1026 @runat_16032006_095252_16710 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_NA_Main:*
                1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
               10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
```

Result No.	Score	Query Match	Length	DB	ID	Description
1	1134	100.0	1713	3	US-09-989-890-105	Sequence 105, App
2	1128	99.5	1785	6	US-10-104-047-799	Sequence 799, App
3	1065.5	94.0	654	6	US-10-074-475-124	Sequence 124, App
4	1002.5	88.4	1977	3	US-09-989-920-56	Sequence 56, Appl
5	1002.5	88.4	1977	9	US-09-989-920-117	Sequence 117, App
6	417	36.8	427	9	US-10-779-543-12051	Sequence 12051, A
7	178.5	15.7	2034	6	US-10-156-761-1440	Sequence 1440, Ap

c	8	178.5	15.7	9025608	6	US-10-156-761-1	Sequence 1, Appli
	9	173	15.3	2914	9	US-10-846-374B-5	Sequence 5, Appli
	10	169.5	14.9	1620	6	US-10-156-761-6900	Sequence 6900, Ap
	11	169.5	14.9	9025608	6	US-10-156-761-1	Sequence 1, Appli
c	12	169	14.9	6567	8	US-10-483-159-27	Sequence 27, Appl
c	13	166	14.6	1799	9	US-10-108-260A-1885	Sequence 1885, Ap
c	14	166	14.6	5142	9	US-10-097-340-42	Sequence 42, Appl
c	15	166	14.6	5142	10	US-11-050-926-42	Sequence 42, Appl
	16	166	14.6	22533	9	US-10-760-493-36	Sequence 36, Appl
c	17	166	14.6	71989	9	US-09-727-889-2	Sequence 2, Appli
c	18	166	14.6	71989	8	US-10-849-462-2	Sequence 2, Appli
	19	166	14.6	164051	9	US-10-760-493-18	Sequence 18, Appl
	20	165.5	14.6	2321	7	US-10-437-963-17115	Sequence 17115, A
c	21	165	14.6	3913	6	US-10-233-045-12	Sequence 12, Appl
c	22	165	14.6	68750	5	US-10-014-717-1	Sequence 1, Appli
c	23	164	14.5	177900	9	US-10-287-436A-1277	Sequence 1277, Ap
c	24	163	14.4	2334	9	US-10-156-761-2734	Sequence 2734, Ap
c	25	162.5	14.3	15231	9	US-09-917-800A-1505	Sequence 1505, Ap
c	26	162.5	14.3	15231	7	US-10-152-319A-1780	Sequence 1780, Ap
	27	162.5	14.3	65140	7	US-10-203-295-1	Sequence 1, Appli
	28	162.5	14.3	125401	7	US-10-203-295-35	Sequence 35, Appl
c	29	162	14.3	23907	5	US-10-077-130-6	Sequence 6, Appli
	30	162	14.3	24120	5	US-10-077-130-4	Sequence 4, Appli
c	31	161.5	14.2	14800	3	US-09-954-456-1601	Sequence 1601, Ap
	32	161.5	14.2	14800	6	US-10-269-909-61	Sequence 61, Appl
c	33	161.5	14.2	14800	7	US-10-717-597-183	Sequence 183, App
c	34	161.5	14.2	14800	9	US-10-843-641A-4628	Sequence 4628, Ap
c	35	161.5	14.2	14800	8	US-10-956-157-418	Sequence 418, App
c	36	161.5	14.2	138203	9	US-10-819-386A-1	Sequence 1, Appli
c	37	160	14.1	2064	7	US-10-437-963-81544	Sequence 81544, A
	38	159.5	14.1	1708	8	US-10-425-115-35257	Sequence 35257, A
c	39	159.5	14.1	4079	6	US-10-305-720-1248	Sequence 1248, Ap
c	40	159.5	14.1	4519	3	US-09-960-706-712	Sequence 712, App
c	41	159.5	14.1	4519	7	US-10-641-643-1202	Sequence 1202, Ap
c	42	159.5	14.1	4519	9	US-10-283-975A-223	Sequence 223, App
c	43	159.5	14.1	14835	6	US-10-240-965-113	Sequence 113, App
	44	159	14.1	753	6	US-10-156-761-7262	Sequence 7262, Ap
	45	158	13.9	3154	9	US-10-846-374B-1	Sequence 1, Appli

```
                          ALIGNMENTS

RESULT 1
US-09-989-890-105
;  Sequence 105, Application US/09999890
;  Publication No. US20040166105A1
;  GENERAL INFORMATION:
;   APPLICANT: Salceda, Susana
;   APPLICANT: Macina, Roberto
;   APPLICANT: Recipon, Herve
;   APPLICANT: Pluta, Jason
;   APPLICANT: Sun, Yongming
;   APPLICANT: Liu, Chenghua
;  TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P.
;  FILE REFERENCE: DEX-0287
;  CURRENT APPLICATION NUMBER: US/09/989,890
;  CURRENT FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: 60/252,509
;  PRIOR FILING DATE: 2000-11-22
;  NUMBER OF SEQ ID NOS: 280
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 105
;  LENGTH: 1713
;  TYPE: DNA
;  ORGANISM: Homo sapien
US-09-989-890-105

Alignment Scores:
Pred. No.:                 2.07e-79              Length:         1713
Score:                     1134.00               Matches:         212
Percent Similarity:        100.0%                Conservative:      0
Best Local Similarity:     100.0%                Mismatches:        0
Query Match:               100.0%                Indels:            0
```

DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-989-890-105 (1-1713)

```
Qy      1 SerProHisGlnAlaAlaAlaProValAspProArgThrLeuAlaThrMetGly   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    104 AGCCCCCACCAAGCCGCACCCGTAGACCAGACCCCCAAGGACCCTGGCCACCATGGGC  163

Qy     21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAla   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    164 CAGAGAGCATTACCTTACCTCTGCTCGCTCTGCTCGCCCGCCCTTGAGTCCCCCACCTGCT  223

Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCGCT  283

Qy     61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyValCysGlyAspAlaAlaProAlaCysLeu   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GCCGGGATTGCCTCCAGCGCTGTGGAGCCTGTGTGCGGGGATGCCAGCCCCTGCCTGTCTA  343

Qy     81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyLeuLysProArgSerThrMetGluCys  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    344 CTGAGGACTCCACTGAGGGACTCTGAGCCAACTGGGCCCAAGGAGCACAATGGAGTGC  403

Qy    101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    404 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAGCAGTCAAGTCAACCA  463

Qy    121 TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    464 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTTCAAAGGCATCCCTGTGTATCCCTACC  523

Qy    141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    524 CGAGGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGATC  583

Qy    161 ProHisProCysAspThrAlaCysSerProAlaProLeuProValValLeuValAlaProArg  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    584 CCCCACCCCATGCGACACAGCGCTGCCAGCACCTTGCCAGTAGTCCTCGTGGCTCCGAGG  643

Qy    181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    644 AGTACTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGGCAGTGGCTCCATGT  703

Qy    201 ArgAlaGluLysLeuMetCysSerSerSerArgSer  212
          ||||||||||||||||||||||||
Db    704 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC  739
```

RESULT 2

US-10-104-047-799

; Sequence 799, Application US/10104047

; Publication No. US20030236392A1

; GENERAL INFORMATION:

;   APPLICANT: HELIX RESEARCH INSTITUTE

;   TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA

;   FILE REFERENCE: H1-A0105

;   CURRENT APPLICATION NUMBER: US/10/104,047

;   CURRENT FILING DATE: 2002-03-25

;   PRIOR APPLICATION NUMBER:

;   PRIOR FILING DATE:

;   NUMBER OF SEQ ID NOS: 4096

;   SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 799

;   LENGTH: 1785

;   TYPE: DNA

;   ORGANISM: Homo sapiens

US-10-104-047-799

Alignment Scores:			
Pred. No.:	3	Length:	1785
Score:	1128.00	Matches:	211
Percent Similarity:	99.5%	Conservative:	0
Best Local Similarity:	99.5%	Mismatches:	1

Query Match: 99.5%   Indels: 0

DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-104-047-799 (1-1785)

```
Qy      1 SerProHisGlnAlaAlaAlaProValAspProArgThrLeuAlaThrMetGly   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 AGCCCCCACCAAGCCGCACCCGTAGACCAGACCCCCAAGGACCCTGGCCACCATGGGC  252

Qy     21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAla   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 CAGAGAGCATTACCTTACCTCTGCTCGCTCTGCTCGCCCGCCCTTGAGTCCCCCACCTGCT  312

Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCGCT  372

Qy     61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyValCysGlyAspAlaAlaProAlaCysLeu   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373 GCCGGGATTGCCTCCAGCGCTGTGGAGCCTGTGTGCGGGGATGCCAGCCCCTGCCTGTCTA  432

Qy     81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyLeuLysProArgSerThrMetGluCys  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433 CTGAGGACTCCACTGAGGGACTCTGAGCCAACTGGGCCCAAGGAGCACAATGGAGTGC  492

Qy    101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    493 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAGCAGTCAAGTCAACCA  552

Qy    121 TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    553 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTTCAAAGGCATCCCTGTGTATCCCTACC  612

Qy    141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    613 CGAGGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGATC  672

Qy    161 ProHisProCysAspThrAlaCysSerProAlaProLeuProValValLeuValAlaProArg  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    673 CCCCACCCCATGCGACACAGCGCTGCCAGCACCTTGCCAGTAGTCCTCGTGGCTCCGAGG  732

Qy    181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    733 AGTACTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGGCAGTGGCTCCATGT  792

Qy    201 ArgAlaGluLysLeuMetCysSerSerSerArgSer  212
          ||||||||||||||||||||||||
Db    793 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAGC  828
```

RESULT 3

US-10-074-475-124

; Sequence 124, Application US/10074475

; Publication No. US20030092898A1

; GENERAL INFORMATION:

;   APPLICANT: Salceda, Susana

;   APPLICANT: Macina, Roberto

;   APPLICANT: Hu, Ping

;   APPLICANT: Recipon, Herve

;   APPLICANT: Karra, Kalpana

;   APPLICANT: Cafferkey, Robert

;   APPLICANT: Sun, Yongming

;   APPLICANT: Liu, Chenghua

;   TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific

;   TITLE OF INVENTION: Genes and Proteins

;   FILE REFERENCE: DEX-0313

;   CURRENT APPLICATION NUMBER: US/10/074,475

;   CURRENT FILING DATE: 2002-02-13

;   PRIOR APPLICATION NUMBER: 60/268,292

;   PRIOR FILING DATE: 2001-02-13

;   NUMBER OF SEQ ID NOS: 295

;   SOFTWARE: PatentIn version 3.1

; SEQ ID NO 124

;   LENGTH: 654

; TYPE: DNA
; ORGANISM: Homo sapien
US-10-074-475-124

Alignment Scores:
Pred. No.:   1.95e-74
Score: 1065.50
Percent Similarity: 95.8%
Best Local Similarity: 95.3%
Query Match: 94.0%
DB:

Length: 654
Matches: 203
Conservative: 1
Mismatches: 8
Indels: 1
Gaps: 1

US-09-989-890-238 (1-212) x US-10-074-475-124 (1-654)

```
Qy      1  SerProHisGlnAlaAlaAlaProValAspThrLeuAlaThrMetGly  20
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db     16  AGCCCCCACCAAGCCGCCGCCGCTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC  75

Qy     21  GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProAla  40
           ||||||||||||||| :|||||||||||||||||||||||||||||
Db     76  CAGAGAGCATTACCTTCATCTCTGCTCCGAGCCGGCCCTTGAGTCCCCACCTGCT  135

Qy     41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136  GCCTGCTCTGGGACCCTGGGTGGGAGTGGTGCCGGCGCCTTCTTGCTTCCGCCGCT  195

Qy     61  AlaGlyIleAlaSerArgAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu  80
           ||||||||| |||||||||||||| |||||||||||||||||||||
Db    196  GCCGGGGATTGCCTCCAGCCGTGCGAGGGGATGCAGCCCCTGCCTGTCTA  255

Qy     81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
           |||||||||||||||||||||||||||||||||||||||||||||
Db    256  CTGAGGGACTCCCCCTGAGGGGACTTGCTGAAGCCAACTGGTTCAAGGAGCACAATGGAGTGC  315

Qy    101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAla---SerGlySerSerGln  119
           |||||||||||||||||| :||||||||||
Db    316  CCCCCAGCCCTGATCGTGCAGCCCCCGGCGGGATGGCCAGGCGGGCTGCAAGTCAA  375

Qy    120  ProTrpAlaAlaAlaSerAlaThrProMetLeuSerLeuSerAlaSerLeuCysIlePro  139
           ||||||||||||||||||||||||||||||||||||||||||||
Db    376  CCATGGGCAGCAGCTTCAGCTACCCCATGTTAAGCTCAAAGGAGCATCCCTGTATCCCT  435

Qy    140  ThrArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpPro  159
           ||||  ||||||||||||||||||||||||||||||||||
Db    436  ACCGAGAGCCACCTCCCCAGCCCCTGATGCGGACTCCTGTCTGCAAGGAGCCACTGGGCCG  495

Qy    160  IleProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaPro  179
           ||||| ||||||||||||||||||||||||||||||||||||||||
Db    496  ATCCCCCACCCCAGCGAGCACAGCCTGCCCGCCACCTTTGCCAGTGTCCTCGTGGCTCCG  555

Qy    180  ArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaPro  199
           ||||||||||||||||||||||||||||||||||||||||||
Db    556  AGGAGTACTATTCTTTCATGAGTCGGACCTGGACATGGCAGTGGCTCCA  615

Qy    200  CysArgAlaGlyLeuMetCysSerSerSerArgSer  212
           ||||||||||||||||||||||||||||
Db    616  TGTCGAGCCGGAAATTGATGTGCTCATCTTCAAGAGC  654
```

; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-989-920-56

Alignment Scores:
Pred. No.:   4.51e-69
Score: 1002.50
Percent Similarity: 94.5%
Best Local Similarity: 93.6%
Query Match: 88.4%
DB:

Length: 1977
Matches: 206
Conservative: 2
Mismatches: 4
Indels: 8
Gaps: 1

US-09-989-890-238 (1-212) x US-09-989-920-56 (1-1977)

```
Qy      1  SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
           ||||||||||||||||||| |||||||||||||||||||||||||||||
Db    357  AGCCCCCACCAAGCCGCCGCCGCCGACCCCTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC  416

Qy     21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla  40
           ||||||||||||| :|||||||||||||||||||||||||||||
Db    417  CAGAGAGCATTACCTTCATCTCTGCTCCGGCCTCTGCTCGTGAGTCCCCACCTGCT  476

Qy     41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla-Al  60
           ||||||||||||||||||||||||||||||||||||||||||||
Db    477  GCCTGCTCTGGGCACCCTGGGTGTGGGAGTGGTGCCGGCGCCTTCTTCGGCCGCCGC  536

Qy     60  aAlaGlyIleAlaSerSerAlaValGlu-ProValCysGlyAspAlaAlaProAlaCysL  80
           |||||||||||  |||||||||||||||||||||||
Db    537  TGCCGGGGATTGCCTCCAGCGCTGTGCGAGGGCTGTGTGCGGGGACGCAGCCCCTGCCTGTC  596

Qy     80  euLeuArgThr-ProLeuArgGlyLeuLeuLysProThrGly-ProArgSerThrMetGl  99
           |||||||||||||||||||||||  ||||||||||
Db    597  TACTGAGGACTCCCACTGAGGGGACTGCTGAAGCCAACTGGTGCCAAGGAGCACAATGGA  656

Qy     99  uCysProProAlaLeuIleValHisProPro-AlaGlyGlyMetAla---SerGlySerS  118
           |||||||||||||||||||||  :::||
Db    657  GTGCCCCCAGCCCTGATCGTGCACCCCCCAGACGCGGCGGGATGGCCAGGCGGGCTGCAA  716

Qy    118  erGlnProTrpAlaAlaAlaSerAlaThrProMetLeuSerLeuSerAlaSerLeuCysI  138
           ||||||||||||||||||||||||||||||||||||
Db    717  GTCAACCATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTA  776

Qy    138  leProThrArg-GlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHis  157
           ||||||||| ||||||||||||||||||||||||||||||
Db    777  TCCCTACCCGAGAGCCACCTCCCCAGCCCCTGATGCGGACTCCTGTCTGCAAGGAGCCAC  836

Qy    158  TrpProIleProHisProCysAsp-ThrAlaCysProAlaProLeuProValLeuVa  177
           |||||||||||||||||| ::::|||||||||||||||||
Db    837  TGGGCCGATCCCCCACCCCAGCGAGCACAGCCTGCCCGCCACCTTTGCCAGTGTCCTCGT  896

Qy    177  lAlaProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaVa  197
           ||||||||||||||||||||||||||||||||||||||||
Db    897  GGCTCCGAGGAGTACTATTCTTTCATGAGTCGGACCTGGACATGGCAGTGGCTCCGAGAAGC  956

Qy    197  lAlaProCysArgAlaGlyLeuMetCysSerSerSerArgSer  212
           |||||||||||||||||||||||||||||
Db    957  GGCTCCATGTCGAGCCGGAGAAATTGATGTGCTCATCTTCAAGAGC  1002
```

RESULT 5
US-09-989-920-117
; Sequence 117, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua

RESULT 4
US-09-989-920-56
; Sequence 56, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500

```
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 117
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-989-920-117

Alignment Scores:
Pred. No.:
Score:                  4.51e-69         Length:          1977
                        1002.50          Matches:         206
Percent Similarity:     94.5%            Conservative:    2
Best Local Similarity:  93.6%            Mismatches:      4
Query Match:            88.4%            Indels:          8
DB:                                      Gaps:            1

US-09-989-890-238 (1-212) x US-09-989-920-117 (1-1977)

Qy    1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  357 AGCCCCCACCAAGCCGCCGCACCCGTAGACCCAGACCCCAAGGACACCCTGGCCACCATGGGC 416

Qy   21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProLeuProProAla 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  417 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGGCCCCTTGAGTCCCCACCTGCT 476

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla-Al 60
          |||||||||||||||||||||||||||||||||||||||||||||||||| :||
Db  477 GCCTGCTCTGGCCGACCCTGGGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCCCGGC 536

Qy   60 aAlaGlyIleAlaSerSerAlaValGlu-ProValCysGlyAlaGlyAspAlaAlaCysL 80
          ||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db  537 TGCCGGGATTGCCTCCAGCGCTGTGGAGCCTGTGCGGAGTGGCAGCCCCTGCCTGTC 596

Qy   80 euLeuArgThr-ProLeuArgGlyLeuLeuLysProThrGly-ProArgSerThrMetGl 99
          |||||||||   ||||||||||||||||||||||||||   |||||||||||||||
Db  597 TACTGAGGACTCCCACTGAGGGGACTGCTGAAGCCAACTGGTGCCAACGGACACAATGA 656

Qy   99 uCysProProAlaLeuIleValHisProPro-AlaGlyMetAla--SerGlySerS 118
          |||||||||||||||||||||||||||   :||||||||   :::
Db  657 GTGCCCCCCAGCCCTGATCGTGCACCCCCAGACCGGGCGGGATGGCCAGGCGGGCTGCAA 716

Qy  118 erGlnProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysI 138
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  717 GTCAACATGGGCGGCAGCACGTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTA 776

Qy  138 leProThrArg-GlyProProProGlnProLeuMetArgThrProAlaAlaArgSerHis 157
          ||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db  777 TCCCTACCCGAGAGGGCCACCTCCCAGCCCTGATGCGGACTCCTGCTGCGAGGAGCCAC 836

Qy  158 TrpProIleProHisProCysAsp-ThrAlaCysProAlaProLeuProValValLeuVa 177
          ||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db  837 TGGCCGGATCCCCCACCCACCCACCGGAGCACAGCCCTGCCCAGCACCTTTGCCCAGTAGTCCTCGT 896

Qy  177 lAlaProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaVa 197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  897 GGCTCCGAGGAGTACTATTCTTTCCATGAGTCCGGACCTGACCTGGGCCCGGAGGATGGGCAGT 956

Qy  197 lAlaProCysArgAlaGlyLysMetCysSerSerSerArgSer 212
          |||||||||||||||||||||||||||||||||||||||||
Db  957 GGCTCCATGTGTGCCGAGCCCGGAGAAATTTGATGTGCTCATCTTCAAGAAGC 1002


RESULT 6
US-10-779-543-12051
; Sequence 12051, Application US/10779543
; Publication No. US20050227917A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12051
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 401
; OTHER INFORMATION: n = A,T,C or G
US-10-779-543-12051

Alignment Scores:
Pred. No.:                          Length:          427
Score:               6.62e-24       Matches:         81
                     417.00         Conservative:    0
Percent Similarity:  100.0%         Mismatches:      0
Best Local Similarity: 100.0%       Indels:          0
Query Match:         36.8%          Gaps:            0
DB:                  9

US-09-989-890-238 (1-212) x US-10-779-543-12051 (1-427)

Qy    1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183 AGCCCCCACCAAGCCGCCGCACCCGTAGACCCAGACCCCAAGGACACCCTGGCCACCATGGGC 242

Qy   21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProLeuProProAla 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  243 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGGCCGGCCCCTTGAGTCCCCACCTGCT 302

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  303 GCCTGCTCTGGCCGACCCTGGGTGGGAGTGGTGCCGGGGCTGCCTTCTGCTTCCGCCGCT 362

Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  363 GCCGGGATTGCCTCCAGCGCTGTGGAGCCTGTGTGCGGNGATGGCAGCCCCTGCCTGTCTA 422

Qy   81 Leu 81
        |||
Db  423 CTG 425


RESULT 7
US-10-156-761-1440
; Sequence 1440, Application US/10156761
; Publication No. US20030119018A1
```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1440
; LENGTH: 2034
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2034)
US-10-156-761-1440

Alignment Scores:
Pred. No.:
Score: 178.50   Length: 2034
Percent Similarity: 37.9%   Matches: 71
Best Local Similarity: 28.6%   Conservative: 23
Query Match: 15.7%   Mismatches: 107
DB: 0.000122   Indels: 47
                                            Gaps: 11

US-09-989-890-238 (1-212) x US-10-156-761-1440 (1-2034)

Qy   2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
                 :::::::::::
Db   839 CCCTGGACGCCCCCCGACCTGGCCGGCCACCGACCTGATGCCGGCGGCCTCAAGG 898

Db 2194 GCCCCTGCCAGTCAGGCTGCCACCAGCTCCTGCACGCCTCCTGCGGTGCT------ 2238
              :::||| |||||||||
Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuLeuSerArgProLeuSerProProAla 40

Db 2239 ------------------GCCAGAGTCTAGCTGCCAGCTGCACATCCTCCCCCTGCC 2286
Qy 2239

Qy   41 ------AlaCysSerGlyAspProGlyCysGlySerGlyAla 52
                 |||||||||||||||||||||
Db 2287 AGCAGGCCTGCTGCGTGCTGCTGCAAGCCAGTCGTGTGTGCCGTCTGCTGCA 2346

Qy   53 GlyLeuProSerAlaSer------ 58
         |||
Db 2347 AGCCTGTCTGCTGCGCAAGCCCATCTGCGTGTGCCGTCTGCTCGGGCTTCCTCTTCAT 2406

Qy   59 AlaAlaAlaAlaGlyIleAlaSerSer------ 76
         |||||::::   AlaValGluCysGlyAspAlaAla
Db 2407 GCTGCCAGCAGTTAGCCGCAGCCGGCTTGCTGCACCACCTCCTGCTGCAGACCCTCCT 2466

Qy   77 ProAlaCysLeuLeuArgProValArgGlyLeuLeuLysProThr------ 92
                                      ----CCCTCCTGCCGCCCCCCG 2493
Db 2467 CCTCCGTGT------

Qy   93 ------GlyProArgSerThrMetGluCysProProAlaLeuIleValHisProProAla 110
                ::: |||
Db 2494 TGTGCAGGTCCACCTGCGTGTGCCATCCGTTCTCCTGTGCCCCTGCCCACCTGCC 2553

Qy   111 GlyGlyMetAlaSerGlySerSerGlnProTrpAlaAlaSerAlaAlaThrProMetLeu 130
                 ||| :::|||
Db 2554 --AGCCCAGGTGCTGCCCGCCGACGTCCCTGCGTCCCTCCTGCCCGCCCA------ 2604

Qy   131 SerSerLysAlaSerLeuCysIleProThrArgGlyProProGlnProLeuMetArg 150
         :::||| ||| GCCAGA 2652
Db 2605 ------CGTGCTCCCGCCTCCTTCCGCGTGCGCCGTCTCCTCAG------

Qy   151 ThrProAlaArgSerHisTrpProIleProHisProCysAspThrAla------ 167
         ::|||||
Db 2653 AGTCCAGTCGCTGACGGGCAGGTCCGGTTCTCCAGAGTTGTTCCAGGGCCATCCTCTGAC 2712

Qy   168 ------CysProAlaProLeuProValValLeuVal 177
                                   :::|||||||||
Db 2713 TTCCGAAAATCCTTGCAGCCCTGAGCCCTGCTGAGTCTCCTCGCCCATAGCCACTGCA 2772

Qy   178 AlaProArgSer------ 181
         |||
Db 2773 CCTCCAGCCACAATGCTGAGCTTTCATGCCGGATTCCCTCCCCAGGCATGCATCTCTCCAC 2832

Qy   182 ThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCysArg 201
         ::||||| CTGGCCAGCCGGTGGTCACTGTTGTCCTGCCCCA 2880
Db 2833 ACCCTCATCTCCATG------

Qy   202 AlaGluLysLeuMetCysSerSerArgSer 212
                  :::|||
Db 2881 TGTGAGTGATCAGTCTGTCAACACAATAAAGTCA 2913

US-09-989-890-238 (1-212) x US-10-156-761-6900 (1-2914)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20

```
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6900
;   LENGTH: 1620
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: (1)..(1620)
US-10-156-761-6900

Alignment Scores:
Pred. No.:              0.0005          Length:         1620
Score:                  169.50          Matches:        75
Percent Similarity:     35.3%           Conservative:   19
Best Local Similarity:  28.2%           Mismatches:     105
Query Match:            14.9%           Indels:         67
DB:                     6               Gaps:           12

US-09-989-890-238 (1-212) x US-10-156-761-6900 (1-1620)

Qy     1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
           :||   |||                              ::|||
Db   557 GCGCCGCGGCTCGGCGCGCCGTCTCCCGGCCCAGCGCACCGAGGCCC---AGCGTGCCGCCC 613

Qy    21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSer--------------ArgProLeuLeuSer 36
           |||  |||  |||             :::::    |||
Db   614 TCGGCCACGCCGCCTCGCTCGGCATCGGTTCCGTCCACGAGTGCGGCGGCCCCCAGATCT 673

Qy    37 ProPro--------------ProAlaAlaCysSerGlyAspPro----------GlyCysGly 49
              |||||          :::|||
Db   674 CCTCCGAGGACGACTTCACCGGCCTGCTCGCCTCGCCCGAGGAGACGGGCCCCCGGG 733

Qy    50 SerGlyAlaGlyLeuProSerAlaSerAlaAlaAlaGlyIle-----------AlaSer 65
           |||                                             :::
Db   734 TCGTCGGCTACTGGGCCGAGTGGGGCGAAGAGAGGCATCGACACGGGCGCGTGCCCTCGGAG 793

Qy    66 SerAlaValGluProValCysGlyAspAlaAla--------------------ProAla 78
           |||:||                          :::                    |||||
Db   794 CGGTCGGCCCGCCGGTGACCTCTTCGTCGACGGCGGCGCTCGGCTCGCACCGCCTGCC 853

Qy    79 CysLeuLeuArgThr-----------------------------------------83
           |||||
Db   854 TGCACCAGCCGTACAGCGACGACGCCGGGCACACGGGTACGGCCCTATCTGGACGACGACGCGG 913

Qy    84 ------------------------ProLeuArgGlyLeuLeuLysProThrGlyProArg 95
                                          |||  :::  |||
Db   914 TCGCCGCCCACGTGATCGCCTGACGGCGGCCTGCTCCCAGCGCGGCGTTCCACGCCATCG 973

Qy    96 SerThrMetGluCysProPro----AlaLeuIleValHisProProAlaGlyValGlyMetAla 114
           |||  |||                 |||  :::
Db   974 TGTGGGGGGCGGCGACGGACGTGCCGTCGGAGGGGCCGCCGCGCCGGAGAAGGTCGGCC 1033

Qy   115 SerGlySerSerGlnProTrpAlaAlaAlaSerAlaThrProMetLeuSer-----------131
           :::|||||                              |||||
Db  1034 TCGCCGCTCGCCGCGCCGCCGCCGGGCCGCTGCAACACGCCGAGATGCTCACCCCCGAGA 1093

Qy   132 -----------------SerLysAlaSerLeuCysIleProThrArgGlyProProPProGln--146
                             |||                 |||                    :::
Db  1094 CCATCGCGGCCTTCGCGGAACTGGCCGCCTCCACGCGGCGTTCCGACGCGGC 1153

Qy   147 ---ProLeuMetGluThrProAlaAlaArgSerHisTrp---------------------158
               |||  |||                            |||
Db  1154 CGGTCGGCGGCGGACGGACGCATGTACGGCGGAGCGGCGGCCCGTACGC 1213

Qy   159 ----------ProIleProHisProCysAspThrAlaCysProLeuProValValLeu 176
                     :::||||                           :::
Db  1214 TCAACCCCTTCGCCGGCGTACCCCCTTGCCTTCGGCTCCGACAGTC 1273

Qy   177 ValAlaProArgSerThr-----IleLeuSerMetSerArgThrTrpThrCysArgArg 194
           ||||||||                           |||
Db  1274 CGGTCACCCGGCTCGAACCGTGGGGCACCGTTCCACGCCGCGGCCTTCCACGCCACGCCGG 1333
```

```
Qy   195 TrpAlaValAlaProCys 200
           |||:||:||::|||
Db  1334 AGCACCGTGTCTCCGTGC 1351

RESULT 11
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;   APPLICANT: OMURA, SATOSHI
;   APPLICANT: IKEDA, HARUO
;   APPLICANT: ISHIKAWA, JUN
;   APPLICANT: HORIKAWA, HIROSHI
;   APPLICANT: SHIBA, TADAYOSHI
;   APPLICANT: SAKAKI, YOSHIYUKI
;   APPLICANT: HATTORI, MASAHIRA
;   TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;   FILE REFERENCE: 249-262
;   CURRENT APPLICATION NUMBER: US/10/156,761
;   CURRENT FILING DATE: 2002-05-29
;   PRIOR APPLICATION NUMBER: JP 2001-204089
;   PRIOR FILING DATE: 2001-05-30
;   PRIOR APPLICATION NUMBER: JP 2001-272697
;   PRIOR FILING DATE: 2001-08-02
;   NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: (4187715)
;     OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.:              1.31            Length:         9025608
Score:                  169.50          Matches:        75
Percent Similarity:     35.3%           Conservative:   19
Best Local Similarity:  28.2%           Mismatches:     105
Query Match:            14.9%           Indels:         67
DB:                     6               Gaps:           12

US-09-989-890-238 (1-212) x US-10-156-761-1 (1-9025608)

Qy     1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
           :||   |||                              ::|||
Db 8258062 GCGCCGCGGCTCGGCGCGCCGTCTCCCGGCCCAGCGCACCGAGGCCC---AGCGTGCCGCCC 8258118

Qy    21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSer--------------ArgProLeuLeuSer 36
           |||  |||  |||             :::::    |||
Db 8258119 TCGGCCACGCCGCCTCGCTCGGCATCGGTTCCGTCCACGAGTGCGGCGGCCCCCAGATCT 8258178

Qy    37 ProPro--------------ProAlaAlaCysSerGlyAspPro----------GlyCysGly 49
              |||||          :::|||
Db 8258179 CCTCCGAGGACGACTTCACCGGCCTGCTCGCCTCGCCCGAGGAGACGGGCCCCCGGG 8258238

Qy    50 SerGlyAlaGlyLeuProSerAlaSerAlaAlaAlaGlyIle-----------AlaSer 65
           |||                                             :::
Db 8258239 TCGTCGGCTACTGGGCCGAGTGGGGCGAAGAGAGGCATCGACACGGGCGCGTGCCCTCGGAG 8258298

Qy    66 SerAlaValGluProValCysGlyAspAlaAla------------------ProAla 78
           |||:||                          :::                  |||
Db 8258299 CGGTCGGCCCGCCGGTGACCTCTTCGTCGACGGCGGCGCTCGGCTCGCACCGCCTGCC 8258358

Qy    79 CysLeuLeuArgThr-----------------------------------------83
           |||||
Db 8258359 TGCACCAGCCGTACAGCGACGACGCCGGGCACACGGGTACGGCCCTATCTGGACGACGACGCGG 8258418

Qy    84 ------------------------ProLeuArgGlyLeuLeuLysProThrGlyProArg 95
                                          |||  :::  |||
Db 8258419 TCGCCGCCCACGTGATCGCCTGACGGCGGCCTGCTCCCAGCGCGGCGGCTTCCACGCCATCG 8258478
```

```
Qy      96 SerThrMetGluCysProPro--AlaLeuIleValHisProProAlaGlyMetAla 114
              ::::|||           :::                ||||||
Db 8258479 GCGACGCCGCCGTGACCTCCGTGTCGAGGGCCGCCGCCGAGAGGTCGGCC 8258538

Qy     115 SerGlySerSerGlnProTrpAlaAlaAlaSerAlaThrProMetLeuSer-------- 131
              ||||||   :::::                 ||||||
Db 8258539 TCGCCCGGTCCCGCGCCCCGGCACCCGTCGAACACGCCGGATGCTCACCCCCGAGA 8258598

Qy     132 --------SerLysAlaSerLeuCysIleProThrArgGlyProProGln-- 146
                   ||||||
Db 8258599 CCATCGGCGGCCTTCGCGGAACTCGGCCTCACCGCCTCGTGCAGCCGCCGC 8258658

Qy     147 --ProLeuMetArgThrProAlaAlaArgSerHisTrp----------- 158
              ||||||
Db 8258659 TGTGGGGGGCGGCAGCGGCCATGTACGGCGACCCTGGGCGCCGAGCGGGCCCGTACGC 8258718

Qy     159 ------ProIleProHisProCysAspThrAlaCysProLeuProValValLeu-- 176
                        :::
Db 8258719 TCAACCCCTTCGCGCCCCTTCCTGCGGGCCGGGCTACCCCTTGCCTTCGGCTCCGACAGTC 8258778

Qy     177 ValAlaProArgSerThr-------IleLeuSerMetSerArgThrTrpThrThrCysArgArg 194
              |||||||
Db 8258779 CCGTCACCCCGCTCGACCCCGTGGGGCACCGTCGCGCGGCGGCCTTCCACCGCACGCCGG 8258838

Qy     195 TrpAlaValAlaProCys 200
              ||||:::||||||
Db 8258839 AGCACCGTGCTCCGTGC 8258856
```

RESULT 12
US-10-483-159-27/c
; Sequence 27, Application US/10483159
; Publication No. US2005003042A1
; GENERAL INFORMATION:
;   APPLICANT: Sabina, Belli
;   APPLICANT: Smith, Nicholas
;   APPLICANT: Wallach, Michael
;   TITLE OF INVENTION: Nucleic Acids Encoding Recombinant 56 and 82 kDa Antigens from
;   TITLE OF INVENTION: Gametocytes of Eimeria Maxima and Their Uses
;   FILE REFERENCE: 65175-A-PCT
;   CURRENT APPLICATION NUMBER: US/10/483,159
;   CURRENT FILING DATE: 2004-01-06
;   NUMBER OF SEQ ID NOS: 42
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
;   LENGTH: 6567
;   TYPE: DNA
;   ORGANISM: Chicken
US-10-483-159-27

Alignment Scores:
Pred. No.:   0.00196   Length: 6567
Score: 169.00   Matches: 56
Percent Similarity: 38.7%   Conservative: 19
Best Local Similarity: 28.9%   Mismatches: 91
Query Match: 14.9%   Indels: 28
DB: Gaps: 8

US-09-989-890-238 (1-212) x US-10-483-159-27 (1-6567)

```
Qy       2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
              ||           :::   ||           |||   ||||||
Db 6416 CCGGAGTACGCGGCAACGGCCA----------CCACCCACTGCTGCCAGCAGTAGTACA 6369

Qy      22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAlaAla 41
              ||   :::   ||||||              ||||||   ::::::
Db 6368 CCACCTCACCCTCCGGCAACAGCTGCCGTGGGGAAGCCGCTGCCCCTTCCTTCTCCTCT 6309

Qy      42 CysSerGlyAspProGlyCysGlySerGlyAlaGlyLeu----------ProSer 56
              ||           ||||||           :::
Db 6308 TCTTCTGCACCGCCTTCGTGGGGTTCCTGTGGAGTTCCGTGAGGGTTGCTCTTCA 6249

Qy      57 AlaSerAlaAlaAlaGlyIleAlaSerSerAlaValCysGlyAspAlaAla 76
              ||||||
Db     ...
```

```
Db 6248 GGTCTCAGCTTCAGGTGTTCCGGTGTCTCAGGGGTTTCCGGCCACTTGCTCG 6189
              :::||||||::|||::::|||           |||               :::
Qy      77 ProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLysProThrGlyThrProArgSer 96
              |||   ||||||
Db 6188 CCTCCTTCACTCGGCCCCTTCTCCCGTCCAGGCTGGACCTCACCACCTTCACCGGCAGCA 6129

Qy      97 ThrMetGluCysProProAlaLeuIleValHisPro--------- 108
              ||||||
Db 6128 CCCTCACCACGGCCGCCTGGCCACCACCCCGTTCCCTGTCTCGCCTTCG 6069

Qy     109 --------ProAlaGlyGlyMetAlaSerGlySerGln-- 119
                       :::
Db 6068 GCTCCTCCAGCACCTGCGCGCCGCCGCCCTTCTCCGGGCAGCTCCTCAGCATTGCCGCCAGAG 6009

Qy     120 ---ProTrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIle 138
                      :::|||
Db 6008 CCTCCTTCTCCGGCGCACCTGGCGCGCCCCTTCTCTCCAGGCAGCTCCTCAGCACTACCG 5949

Qy     139 ProThrArgGlyProProGlnProLeu--MetArgThrProAlaAlaArgSerHis 157
              ||||||
Db 5948 CCAGAGCCTCCTTCCGCCGGCACCTGCGCGCCGCCGCCCTTCTCCGGGCAGCTCCTCCAGCA 5889

Qy     158 TrpProIleProHisProCysAspThrAlaCysProAlaPro 171
              ||||||:::
Db 5888 CTACCGCCAGAGCCTTCGCCAGCACCCCCTTCTCCG 5847
```

RESULT 13
US-10-108-260A-1885/c
; Sequence 1885, Application US/10108260A
; Publication No. US2004005560A1
; GENERAL INFORMATION:
;   APPLICANT: HELIX RESEARCH INSTITUTE
;   TITLE OF INVENTION: H1-A0106
;   FILE REFERENCE: No. US20040005560A1e1 full length cDNA
;   CURRENT APPLICATION NUMBER: US/10/108,260A
;   CURRENT FILING DATE: 2002-03-27
;   NUMBER OF SEQ ID NOS: 5458
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1885
;   LENGTH: 1799
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-1885

Alignment Scores:
Pred. No.:   0.00103   Length: 1799
Score: 166.00   Matches: 65
Percent Similarity: 42.0%   Conservative: 19
Best Local Similarity: 32.5%   Mismatches: 93
Query Match: 14.6%   Indels: 23
DB: Gaps: 5

US-09-989-890-238 (1-212) x US-10-108-260A-1885 (1-1799)

```
Qy       1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuSerArgProLeuSerPro----------- 37
              |||||||           :::   ::                 :::|||         ::|||
Db 1455 TCTCCTGCATCTTCCTTTCTGATCTGATCTCATATCCTCCTTCCTTCTCCTCTCCTGGCCCCCCACATCTTCT 1396

Qy      21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerPro--------- 37
              ||||||                        ||||||:::
Db 1395 CCTGCCCCCACATCTTCTCCTCCTCCTCCGCCCCCCACATCTTCTCCTTGGCCCCTGGCCCCAGATCTTCT 1336

Qy      38 --------ProProAlaAlaCysSerGlyAspProGlyCysGlySer 50
                       ||||||
Db 1335 CCTCCTGGCCCCCGCATCTTCTCCTCCTGTTCCCGGATCTTCTCCCCGGCCCCGCATCT 1276

Qy      51 GlyAlaGlyLeuProSerAlaSerAlaAlaAlaAlaGlyIleAlaSerSerAlaValGluPro 70
              ||||||:::
Db 1275 TCTCCTCCTGCTCCC---GCATCTTCTCCTCCTGCCCCTGCCCCTCCTCCTGTCCCC 1219

Qy      71 ValCysGlyAspAlaAla-CysLeuLeuArgThrProLeuArgGlyLeuLeuLysGly 90
              ||||||
Db     ...
```

```
Percent Similarity:     36.0%        Conservative:    14
Best Local Similarity:  29.8%        Mismatches:      56
Query Match:            14.6%        Indels:          90
DB:                     5            Gaps:            12

US-09-989-890-238 (1-212) x US-10-097-340-42 (1-5142)

Qy    19 MetGlyGlnArgAlaLeuProSer------LeuAlaLeuSerArgProLeu--  35
            :::||||     :::||||||||       ||::|||
Db  2006 CTCGGGCCTGCTCATTCCCTGCAAGACCTTATCTTCCTCTCCAAGTTCCTCTTCCA 1947

Qy    36 SerProPro---ProAlaAlaCysSerGly-----------AspProGly  47
         ||||||||     |||  ::||:|||
Db  1946 TCTCCTCCTTCTCCATTCGGAGCTGCAGGAGTTCCTGCTGGTGGTGGCTCAAGATCCTCCT 1887

Qy    48 CysGlySerGlyAla-------------------------GlyLeuPro  55
            |||||||  |||                        |||
Db  1886 TGTTCTTCTGGAACATACTCTGCCAATGCCCGTCCTCCTGAAGTCTCCTCCAGCTCCC 1827

Qy    56 SerAlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluPro--------  70
         ||||||::|  |||   ||||||||
Db  1826 TCTGCAGCCCTGCACAAGGCTGACATCTTCTGCCTCTCGGCTCTCCAGCACTGCATGGT 1767

Qy    71 ---------ValCysGlyAspAlaAlaProAlaCysLeuLeuArgThrProLeu  85
                     ||     |:||:||||
Db  1766 CCTGGGTTGCATCTTGCATGCTCCTGCGGGAGCTGCTCGTGTCTCGCGTGGTACT---GCT 1710

Qy    86 ArgGlyLeuLysProThrGlyProArgSerThrMetGluCysProProAlaLeuIle 105
         |||||||||
Db  1709 GCCGGACATGTTCCACCTCCTGGT----------------------------------- 1686

Qy   106 ValHis-------ProProAlaGlyMetAlaSerGlySerGlnProTrpAla 122
         |||||||     ||||||
Db  1685 --CACGGGAGGTACCTCCTCTTTCAGGGCCCCCTTCAGGGCTGTCAACTCCCGCTCCC 1629

Qy   123 AlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuProThrArgGly 142
         ||||||||  ::::|||::|:
Db  1628 GCAGCCTCCAGCTGCTGCTCCTCTACTCGCTGTTTCC----------------- 1596

Qy   143 ProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIleProHis 162
         ||     |||:|||
Db  1595 ---CCTTCAAGACCTCTTCCAGAAGTTCCCGGGTCT---------------------- 1563

Qy   163 ProCysAspThrArgCysProAlaProLeuProValValLeuValAlaProArgSerThr 182
              |||||  |||||
Db  1562 -----------------------------------CTAACAGGTCCTTCAGTAACA 1542

Qy   183 -IleLeuSerMetSerArgThr-TrpThrCys---------------------ArgArgT 195
          ::||||||    ||||||||
Db  1541 CCTCCTTAGCAGGCTCAGGACCCTGGACATGTTTCAGCTTGTTCGCAGCTCCATCACCT 1482

Qy   195 rpAlaValAlaProCysArg 201
         :::|||||||||
Db  1481 GGGTCTCCAGCCCATGTCGT 1462
```

```
Db  1218 GCATCTTCTCCTCCTGCTTCCTCCTCCGATCTTCTCCTCCTGGCCCCGCATCT-------TC 1168
            |||
Qy    90 sProThrGlyProArgSerThrMetGluCysProProAlaLeuIleValHisProProAl 110
         |||   |||||::   ||||||||||||
Db  1167 TCCTCCTGTCCCCCGCATCTTCTCCTCCTGCCCCCGCATCTTCTCCTCCTGCCCCCGC 1108

Qy   110 aGlyGlyMetAlaSerGlySerGlnProTrpAlaAlaSerAlaThrProMetLe 130
            :::   |||||||    ::|||  |||
Db  1107 ATCTCCTCCTCCTGCCCCCGCATCTTCTCCTCCTGGCCCCGCATCTTCTCCTCCT----- 1053

Qy   130 uSerSerLysAlaSer-LeuCysIleProThrArgGlyProProGlnProLeuMetA 150
            :::   |||||||||     ::|
Db  1052 ----GCCCCGCATCTTCTCCTGCCCCCGCATCTTCTCCTGCCCCCGCCCCACATCTT 997

Qy   150 rgThrProAlaAlaArgSerHisProIlePro-------HisProCysAspThrA 167
            ||||||||   ||||||
Db   996 CTTCCTCCTGCCCCCACATCTTCTCCTCCTGGCCCCGCCCACATCTTCTCCTCCTGCCCCCACAT 937

Qy   167 laCysPro-AlaProLeuProValValLeuValAlaProArgSerThrIleLeu 184
            ::    ||   :::::|||||
Db   936 CTTCTCCTCCTGCCCCGCCCCACATCTTCTCCTCCTGCCCCCGCCCCATCTTCTCCTCCTG 883
```

```
Alignment Scores:
Pred. No.:        0.00269       Length:      5142
Score:            166.00        Matches:     68
```

```
;  APPLICANT:  Rosemarie SCHMANDT
;  APPLICANT:  Xumei ZHAO
;  APPLICANT:  Karen GLATT
;  TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
;  TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
;  FILE REFERENCE: MRI-030
;  CURRENT APPLICATION NUMBER: US/11/050,926
;  CURRENT FILING DATE: 2005-02-04
;  PRIOR APPLICATION NUMBER: US/10/097,340
;  PRIOR FILING DATE: 2002-03-14
;  PRIOR APPLICATION NUMBER: 60/276,025
;  PRIOR FILING DATE: 2001-03-14
;  PRIOR APPLICATION NUMBER: 60/325,149
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: 60/276,026
;  PRIOR FILING DATE: 2001-03-14
;  PRIOR APPLICATION NUMBER: 60/324,967
;  PRIOR FILING DATE: 2001/09/26
;  PRIOR APPLICATION NUMBER: 60/311,732
;  PRIOR FILING DATE: 2001-08-10
;  PRIOR APPLICATION NUMBER: 60/325,102
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: 60/323,580
;  PRIOR FILING DATE: 2001-09-19
;  NUMBER OF SEQ ID NOS: 363
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 42
;    LENGTH: 5142
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-11-050-926-42

Alignment Scores:
Pred. No.:              0.00269          Length:          5142
Score:                  166.00           Matches:         68
Percent Similarity:     36.0%            Conservative:    14
Best Local Similarity:  29.8%            Mismatches:      56
Query Match:            14.6%            Indels:          90
DB:                     10               Gaps:            12

US-09-989-890-238 (1-212) x US-11-050-926-42 (1-5142)

Qy     19 MetGlyGlnArgAlaLeuProSerSer-----LeuAlaLeuSerArgProLeu--- 35
             :::|||||||||||     ||||:::   |||:|||  |||||::::|
Db   2006 CTCGGGCCTGCTCTAATTCCCTCGCAAGACCTCTATCTTCTCCAAGCTCCTCTTCCA 1947

Qy     36 SerProPro---ProAlaAlaCysSerGly---------------AspProGly 47
             |||||||||   |||  ||||  |||:::              ||||||
Db   1946 TCTCCCTCCTTCTCCATTCGCAGCTGCAGGAGTTCCTGCTTGGGCTCTAAGATCCTCCT 1887

Qy     48 CysGlySerGlyAla---------------------------------GlyLeuPro 55
             ||||||||||                                      |||
Db   1886 TGTTCTTCTGGAACATACTCTGCCAATGCCCTGTCTCCCTCGAAGTCTCCTCCAGCTCCC 1827

Qy     56 SerAlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluPro---------- 70
             |||||:::  ||| |||||||||||||||
Db   1826 TCTGCAGCCCTGCACAAGGGCTGACATCTTCTGCCTTCCGGCCTCCAGCACTGCATGGT 1767

Qy     71 ----------ValCysGlyAspAlaAlaProAlaCysLeuLeuArgThrProLeu 85
                       |||  |||    ||| ||||||:::||||||
Db   1766 CCTGGGTTGCATCTGCATCCTTGCCTCGCCGGAGCTGCCTCGTGTCTCGGTGGTACT---GCT 1710

Qy     86 ArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCysProProAlaLeuIle 105
             ||     :::          |||||||||||||
Db   1709 GCCGGACATGTTCCACCTCCTGGT-------------------------------------- 1686

Qy    106 ValHis--------ProProAlaGlyGlyMetAlaSerGlySerSerGlnProTrpAla 122
             ||||:         |||||||     ||||||||
Db   1685 ---CACCGGAGGCTACCTCCTCTTTCAGGGCCCCCCTTCAGGGGTGTCAACTCCCGCTCCC 1629

Qy    123 AlaAlaSerAlaThrProMetLeuSerLysAlaSerLeuCysIleProThrArgGly 142
             ||||||||||     ||||:::||
Db   1628 GCAGCCTCCAGCTGCTCCTCCACTCGCTGTTTCC-------------------------- 1596
```

```
Qy    143 ProProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIleProHis 162
             |||||||||  ||::::||||||:::
Db   1595 ---CCTCCAAGACCTCTTCCAGAAGTTCCCGGGTCT--------------------------- 1563

Qy    163 ProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArgSerThr 182
             |||  |||  |||||||||  |||     |||     ||
Db   1562 --------------------------------CTAACAGGTCCTTCAGTAACA 1542

Qy    183 -IleLeuSerMetSerArgThr-TrpThrCys---------ArgArgT 195
             :::||||||  |||||||||
Db   1541 CCTCCTTAGCAGGCTCAGGACCCTGGACATGTTTCAGCTTGTTCTGCAGCTCCATCACCT 1482

Qy    195 rpAlaValAlaProCysArg 201
             :::|||||||||
Db   1481 GGGTCTCCAGCCCATGTCGT 1462
```

Search completed: March 17, 2006, 07:16:30
Job time : 4711 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 17, 2006, 05:46:32 ; Search time 1473 Seconds
(without alignments)
335.588 Million cell updates/sec

Title: US-09-989-890-238
Perfect score: 1134
Sequence: 1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop   6.0 , Delext   7.0

Searched: 8023312 seqs, 1165852854 residues

Total number of hits satisfying chosen parameters: 16046624

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095254_16764/app_query.fasta_1
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCU=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abss03h
-USER=US09989890 @CGN_1_1_431 @runat_16032006_095254_16764_NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Published Applications NA_New:*
1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1128	99.5	1785	9	US-11-072-512-799	Sequence 799, App
2	163.5	14.4	1306	12	US-11-096-568A-9547	Sequence 9547, Ap
c 3	162.5	14.3	15234	12	US-11-136-527-306	Sequence 306, App
4	162	14.3	2052	9	US-11-096-568A-16769	Sequence 16769, A

	Score	% Match	Length	DB	ID	Description
c	5	162	14.3	23907	12	US-11-186-731-6	Sequence 6, Appli
c	6	162	14.3	24120	12	US-11-186-731-4	Sequence 4, Appli
	7	158	13.9	1590	8	US-10-858-730-162	Sequence 162, App
	8	157	13.8	2625	8	US-10-858-730-36	Sequence 36, Appl
	9	157	13.8	3334	12	US-10-330-773-3025	Sequence 3025, Ap
c	10	156.5	13.8	127722	7	US-10-330-773-278	Sequence 278, App
c	11	155.5	13.7	14619	12	US-11-128-061-745	Sequence 745, App
c	12	155.5	13.7	14619	12	US-11-128-049-745	Sequence 745, App
c	13	154.5	13.6	1835	9	US-11-096-568A-23836	Sequence 23836, A
	14	154	13.6	1269	9	US-11-096-568A-21912	Sequence 21912, A
	15	154	13.6	1676	9	US-11-096-568A-19912	Sequence 19912, A
	16	153.5	13.5	1702	9	US-11-096-568A-12594	Sequence 12594, A
c	17	152	13.4	2093	9	US-11-072-512-1666	Sequence 1666, Ap
c	18	151.5	13.4	1767	9	US-11-096-568A-12181	Sequence 12181, A
	19	150.5	13.3	1473	9	US-11-110-082-8	Sequence 8, Appli
	20	149.5	13.2	2740	9	US-11-096-568A-24025	Sequence 24025, A
	21	149	13.1	1277	12	US-11-113-424-5	Sequence 5, Appli
	22	149	13.1	1322	12	US-11-113-424-7	Sequence 7, Appli
	23	149	13.1	2514	12	US-11-052-554A-543	Sequence 543, App
	24	149	13.1	355211	7	US-10-330-773-242	Sequence 242, App
c	25	148.5	13.1	1403	9	US-11-096-568A-18205	Sequence 18205, A
	26	148.5	13.1	2010	12	US-11-058-727-5	Sequence 5, Appli
	27	148.5	13.1	2010	12	US-11-108-389-5	Sequence 5, Appli
	28	148.5	13.1	2010	12	US-11-224-624-5	Sequence 5, Appli
	29	148.5	13.1	3624	7	US-10-755-092-6	Sequence 6, Appli
c	30	148.5	13.1	11736	12	US-11-000-463-218	Sequence 218, App
	31	147	13.0	1756	9	US-11-096-568A-15790	Sequence 15790, A
c	32	146.5	12.9	911	8	US-10-750-185-44656	Sequence 44656, A
c	33	146.5	12.9	911	8	US-10-750-623-44656	Sequence 44656, A
	34	146	12.9	1282	9	US-10-750-623-15585	Sequence 15585, A
c	35	146	12.9	1844	8	US-10-973-115B-475	Sequence 475, App
	36	146	12.9	1844	7	US-10-131-826A-475	Sequence 475, App
c	37	145	12.8	1820	9	US-11-072-512-1089	Sequence 1089, Ap
c	38	145	12.8	2406	12	US-11-052-554A-550	Sequence 550, App
	39	145	12.8	121736	9	US-11-114-798-49	Sequence 49, Appl
c	40	145	12.8	318488	9	US-11-114-798-58	Sequence 58, Appl
	41	144.5	12.7	96256	8	US-10-775-169-352	Sequence 352, App
	42	144	12.7	2965	8	US-10-750-185-36860	Sequence 36860, A
	43	144	12.7	2965	8	US-10-750-623-36860	Sequence 36860, A
	44	144	12.7	191684	12	US-11-121-086-2	Sequence 2, Appli
	45	143.5	12.7	1328	9	US-11-096-568A-8975	Sequence 8975, Ap

ALIGNMENTS

RESULT 1
US-11-072-512-799
; Sequence 799, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978

```
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 799
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-072-512-799

Alignment Scores:
Pred. No.:
Score:                          1.13e-47
Best Local Similarity:          1128.00
Percent Similarity:             99.5%        Length:         1785
                                99.5%        Matches:        211
Query Match:                    99.5%        Conservative:   0
DB:                             9            Mismatches:     1
                                             Indels:         0
                                             Gaps:           0

US-09-989-890-238 (1-212) x US-11-072-512-799 (1-1785)

Qy    1   SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193   AGCCCCCACCAGCCGCACCCGTAGACCCAGACCCCAAGGACCCCTGGCCACCATGGGC     252

Qy   21   GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla  40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  253   CAGAGAGAGCATTACCTTCATTCTCTGGCTCTTCTGGCTCGTGAGCGCCGGCCGCCGCCT  312

Qy   41   AlaCysSerGlyAlaAspProGlyCysGlyAlaGlyLeuProSerGlyAlaSerAlaAla  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  313   GCCTGCTCTGGCGCACCCTGGGTGTGGAGTGGTGCCGGGCTCCTTCGTTCTCCGCCGCT   372

Qy   61   AlaGlyIleAlaSerSerAlaValGluProValCysGlyAlaAspAlaAlaAlaProAlaCysLeu  80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  373   GCCGGGATTGCCTCCAGCGCTGGAGCCGTGTGCGGGGATGCAGCCCCTGCCTGTCTA    432

Qy   81   LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  433   CTGAGGACTCCACTGGAGGGGACTCGTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC  492

Qy  101   ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  493   CCCCCAGCCCTGATCGTGCACCCCCAGCCGGCGGGGATGGCCAGCTCAAGTCAACCA    552

Qy  121   TrpAlaAlaAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  553   TGGGCAGCAGCTTCAGCTACCCCGATGTGTAAGCTCAAAGGCATCCCTGTTATCCCTACC  612

Qy  141   ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  613   CGAGGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCAAGGAGCCCACTGGCCGATC  672

Qy  161   ProHisProCysAspThrAlaCysProLeuProLeuValValAlaValAlaProArg  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  673   CCCCACCCATGCGACACAACGCTGCCCAGCACCTTTGCCCAGTCCTCCTGGCTCCGAGG   732

Qy  181   SerThrIleLeuSerMetSerSerArgThrThrCysArgArgTrpAlaValAlaProCys  200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  733   AGTACTATTCTTTCCATGAGTCGGACCTGCACCTGGACGATGGGCAGTGGGCTTCCATGT  792

Qy  201   ArgAlaGluLysMetLeuMetCysSerSerSerArgSer  212
          |||||||||||||||||||||||||||||||||||||||
Db  793   CGAGCCGAGAAAATTGATGTGCTCATCTTCAAGAAGAGC    828

RESULT 2
US-11-096-568A-9547
; Sequence 9547, Application US/11095568A
; Publication No. US2006004824OA1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
```

```
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 9547
; LENGTH: 1306
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1306)
; OTHER INFORMATION: Ceres Seq. ID no. 12613953
US-11-096-568A-9547

Alignment Scores:
Pred. No.:                      6.93
Score:                          163.50
Percent Similarity:             38.4%        Length:         1306
Best Local Similarity:          28.8%        Matches:        72
Query Match:                    14.4%        Conservative:   24
DB:                             9            Mismatches:     94
                                             Indels:         60
                                             Gaps:           11

US-09-989-890-238 (1-212) x US-11-096-568A-9547 (1-1306)

Qy    6   AlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGlnArgAlaLeu-  24
          ||||:||::||||      ::::::      |||
Db  187   GCCGCGCCGATCGACAACTCGCCCTACATCGTGTCCGGGCTCCCGGACAAGTCGCTGCTG  246

Qy   25   ------------ProSer-SerLeuAlaLeuLeuSerArgProLeuSer  36
                       |||:|||
Db  247   GTCTGGACCTCCACCAACCCGGTGCCGTCCCTCGGCGACGGCACCGTCGGCACCGACTAC  306

Qy   36   rProProAlaAlaCys---------------------------SerGlyAb  45
          |||
Db  307   GCCGTGCCCTTCCGCGCCCTCACCGGCCACTTCGTCCAGGACGGTGGTCATCAGC     366

Qy   45   pProGlyCysGlySerGlyAlaGlyLeuPro---SerAlaSerAlaAlaAlaGlyIleAl  64
          :::             ::::::||:
Db  367   TCCGACGGCCAGTTCGCGCTTCCGGGACGGGCGGAGCGTCCCGCCTCCGGACCTC   426

Qy   64   aSerSerAlaValGluProValCysGlyAspAlaAlaAlaProAlaCysLeuLeuArgThrPr  84
          |||                 ||||||:
Db  427   TCCACCGGCCTCACCACCCGCGCGTTCGTCGGCCACCACCAAGGACGTCATCTCGGTCGCC  486

Qy   84   oLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCysProProAlaLe  104
                                 |||||||
Db  487   T--------------------------TCTCCGTCGACAACCGCCGATCGTCTCCGCCTCC  522

Qy  104   uIleValHisProProAlaGlyGlyMetAlaSerGlySerGln----------------  119
                                   |||       :::
Db  523   CGGCGACAACCATCAAGCTCTGGAACACCCTCGGCGAGTGCAAGTACACCATCGGCGGC  582

Qy  120   -ProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIlePr  139
           |||::||||      :||:||||:|
Db  583   GACATGGGCGGCGGCGAGGGCCACCACCG----GCTGGGTCGTGCGTCAGGTTCTCCCCC  639

Qy  139   oThrArgGlyProProGln----------------ProLeuMetArg----------  150
           |||       ::::|||
Db  640   AACATCCAGCAGCCCACCATCGTCTCCGGCTCCTGGGACAAGACCGTCAAGGTCTGGAAT  699

Qy  151   -ThrProAlaAlaArgSerHisTrpProIleProHisProCysAspThrAlaCysProAl  170
           :::|||           |||:::
Db  700   CTCACCAACTGCAAGCTCCGCTCCACCCCTCGCCGTGGTGTACGGTGAACGCCGTC  759

Qy  170   aPro---LeuProValValLeuValAlaValAlaProArgSerThrIleLeuSerMetSerSerArgTh  189
           :||:           |||:::
Db  760   GCCGTGAGCCCGATGGCTCAGCCCTCCTGCGGCCTCCGCGGCCAAGGACGGCGGCGTCACGCTGCTC  819

Qy  189   rTrpThrCysArgArg-------------------TrpAlaValAlaProCysArgAlaGl  203
           |||||  ||||
Db  820   TGGGACCTGTCCGAGGGGGAAGAGGCTGTACCAGCTGGAAGCGCGGGCTCCA-------  868

Qy  203   uLysLeuMetCysSerSerSerArgSer  212
```

Db　　　869　------------TCATCCATGCGCGCTCT　883

RESULT 3
US-11-136-527-306/c
; Sequence 306, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
;　APPLICANT: Wyeth
;　APPLICANT: Mounts, William M
;　TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
;　FILE REFERENCE: 031896-041000 (AM101086)
;　CURRENT APPLICATION NUMBER: US/11/136,527
;　CURRENT FILING DATE: 2005-05-25
;　PRIOR APPLICATION NUMBER: US 60/574,294
;　PRIOR FILING DATE: 2005-05-26
;　SOFTWARE: PatentIn version 3.2
;　SEQ ID NO 306
;　LENGTH: 15234
;　TYPE: DNA
;　ORGANISM: Rattus norvegicus
US-11-136-527-306

Alignment Scores:
Pred. No.:　　　　　　39.6　　　　　　　Length:　　　　　15234
Score:　　　　　　　　162.50　　　　　　Matches:　　　　　76
Percent Similarity:　39.2%　　　　　　Conservative:　　29
Best Local Similarity:　28.4%　　　　Mismatches:　　　80
Query Match:　　　　　14.3%　　　　　　Indels:　　　　　83
DB:　　　　　　　　　　12　　　　　　　　Gaps:　　　　　　13

US-09-989-890-238 (1-212) x US-11-136-527-306 (1-15234)

Qy　　13　ProArgThrLeuAlaThrMetGlyGlnArgAlaLeuPro------------Ser　26
　　　　　　　||||||||||||　　：：：　　　||||||　　　　　　：：：
Db　7141　CCAGCTCCTCCATCTGCACGCGGACAGAGAAGAGCTCCTCCTTACCTGGCTACGCTGCC　7082

Qy　　27　SerLeuAlaLeuLeuSerArgProLeuSerProProAlaAlaCysSerGlyAspPro　46
　　　　　　|||||　　||||||　||||||||　　：：：||||||　　　　　　：：：　||
Db　7081　GGGGTCGCCCTGTTACCTCAGCCTTTA------------GCCGCTCGAGCTCCTCATCCA　7034

Qy　　47　GlyCysGlySerGlyAlaGlyLeuLeuProSerAlaSerAlaAlaAlaGlyIleAla------　64
　　　　　　|||||　||||||　　|||||||||||：||||　　|||||：：：||||||
Db　7033　GGATGCTCTTCTGGGTGGTCGGTCTCTCGAGCTGCAGCGCTGGTCAGCTCCTGCT　6974

Qy　　65　------------SerSerAlaValGluProValCysGlyAspAlaAlaPro　77
　　　　　　：：：||||||||||：：||||：：：||||
Db　6973　CTACCTGAGCCTTCTGCCGTAGCGTCTGCTCTGCGCAAACTTCTTGTGCTTCCATCTCCG　6914

Qy　　78　AlaCysLeuLeuArgThrProLeuArgGlyLeuLysProGlyLeuLysProGlySerThr　97
　　　　　　|||　||||||　||||||||　：：：||||||||
Db　6913　CGT------------CGGCTGCCTGCTCGTTCAACGCAGCCCAGCTCCGCTGGGCCC--　6866

Qy　　98　MetGluCysProPro------------AlaLeuIleValHisProProAlaGly　111
　　　　　　：：：||　||||||　　：：：||||||||||
Db　6865　--GACGCGCGCCCTTCCTGCGTCCGCTTCCTGCGAGCCCCTGCGCCT　6809

Qy　112　GlyMetAlaSerGlySerSerGlnProTrpAlaAlaSerAlaSerAlaThrProMet---Leu　130
　　　　　　||||||||||　　||||||||||||||　||||||
Db　6808　GGGCCTGGGCCTGAGCCTGCTCCTCTCCGACTGCTTCAGCCGCTGGCCTTCTCCACTT　6749

Qy　131　SerSerLysAlaSerLeu------CysIleCysProThr---　146
　　　　　　：：：||||　：：：　|||||||
Db　6748　GCTTCCTAGACTGCCGCTGCCCAGTTCGTCCTGCTCCGCTCCCGCCTCCTCCAG　6689

Qy　147　------------ProLeuMetArgThr------------　151
　　　　　　||||||||||||
Db　6688　CAGGTCGCCCCGCCTGCCCCGCTCCTCCGCAGCCGCTCCAGCATGTTCTGCTCTTGCT　6629

Qy　151　------------------　151
　　　　　　：：：||||||：：：：
Db　6628　GAAGAGTCGCTGCTGCGAGCTCCTTCTTCGCTGCTGCCACCACAAAGGCGTGCGCCTTCTCCT　6569

Qy　152　---ProAlaAlaArgSerHisTrpProIleProHisProCysAspThrAlaCysProAla　170
　　　　　　||||||||||
Db　6568　CCGGCTGCAGCCGTTCTGGCCAGCAGCCT---CCTGGGCGCCAGTGCAGCTGCCGCAGACT　6512

Qy　171　ProLeuProValValLeuVal-AlaProArg-SerThrIleLeuSerMetSerArg---　188
　　　　　　：：：||||||||||　||||||||||||：：：：：||||||
Db　6511　CCTGCTCAGCTCGCTCCTCCGCAGGGCGCCGCTTCCTCCAACCTTGGCCTTGAGCCGCTCGA　6452

Qy　189　------------ThrTrpThrCysArgArgTrpAlaValAlaProCysArgArgAlaGluL　204
　　　　　　：：：
Db　6451　CTTCCTCCAGTGCAGCACTTGCGCTGCCGT------------GCCGCTTCCTCCTCCGCTGCCA　6401

Qy　204　ysLeuMetCysSerSerSer　210
　　　　　　：：：|||：：|||：||
Db　6400　GGCTCCTCTGCACCCGCTCC　6381

RESULT 4
US-11-096-568A-16769
; Sequence 16769, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
;　APPLICANT: Alexandrov, Nickolai et al.
;　TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
;　TITLE OF INVENTION: Therby
;　FILE REFERENCE: 2750-1592PUS2
;　CURRENT APPLICATION NUMBER: US/11/096,568A
;　CURRENT FILING DATE: 2005-04-01
;　NUMBER OF SEQ ID NOS: 34471
;　SEQ ID NO 16769
;　LENGTH: 2052
;　TYPE: DNA
;　ORGANISM: Zea mays subsp. mays
;　FEATURE:
;　NAME/KEY: misc_feature
;　LOCATION: (1)..(2052)
;　OTHER INFORMATION: Ceres Seq. ID no. 12354033
US-11-096-568A-16769

Alignment Scores:
Pred. No.:　　　　　　11.1　　　　　　　Length:　　　　　2052
Score:　　　　　　　　162.00　　　　　　Matches:　　　　　74
Percent Similarity:　34.0%　　　　　　Conservative:　　15
Best Local Similarity:　28.2%　　　　Mismatches:　　　95
Query Match:　　　　　14.3%　　　　　　Indels:　　　　　78
DB:　　　　　　　　　　9　　　　　　　　Gaps:　　　　　　13

US-09-989-890-238 (1-212) x US-11-096-568A-16769 (1-2052)

Qy　　7　AlaProValAspGlnThrProArgThrLeuAlaThrMet------------GlyGlnArg　22
　　　　　||||||||||：：：　||||||||||
Db　635　GCTCCCTCCCCTCCCCAGGCGCGGAGCAGTTCGACGAGCATGCGGCGGGGGCGCAAGG　694

Qy　23　AlaLeuPro------------SerSerLeuAlaLeuLeuSerArgProLeuSerPro　37
　　　　　||||||　　　：：：||||||||||
Db　695　GCCGCTCCACCTTCTTCAACTACTTCGTCTTCGCCTCCTCGCCCCTGCTCGCCCG　754

Qy　38　ProProAlaAlaCysSerGlyAsp------------ProGlyCysGlySerGlyAlaGly　53
　　　　　||||：：：　　　：：：||||||||||
Db　755　TCACCTTGCCG---TCTGGGTGCGAGGACAACAAGGGCTGGCAGTGGGGTTCGGCATCT　811

Qy　54　LeuProSerAlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluProValCysGly　73
　　　　　||||||||||||||||||
Db　812　CCACCATGGCCATCCTGCTCTCCATCCCGCCTTCGTCGCCGGCTCCAGCCGCTACCGGA　871

Qy　74　AspAlaAlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLysProThrGly　93
　　　　　|||
Db　872　ACAAGGTGCCCA------------CGGGGAGCCCGGCGTCACCACCATCG　907

Qy　94　ProArgSerThrMetGluCysProProAlaLeuIleValHisProProAlaGlyGlyMet　113
　　　　　|||
Db　908　CCAAGGTCCTCA------------TCGCCGGCGCCGTGCGCGGCGCC　940

```
Qy   40 AlaAla----------------------------------CysSerGlyTrpAspProGlyCysGlySerGlyAlaGlyLeu  54
           |||||                                          ||||||||||||||||||||||||||||||||
Db 4507 GCTGCACCACCAGCTCCGTGCGAACGGCGCCGTCATCGACCGCGCGGCCCAGCCTCCATGCGCACTT-----------  4460

Qy   55 ProSerAlaSerAlaAlaAlaGlyIleAlaSerAlaValGluProValCysGly---  73
           ||||:::     ||||:::  |||
Db 4459 ---TCGAGCTGAAGCTCAGCTTCTTCCCGTCCTTGTACCACATCACCTCCGTCTGGGCCT  4403

Qy   74 ----------------------------------AspAlaAlaProAlaCysLeu  80
                                              ||:::
Db 4402 GAGCCACCTCGCAGCTGAGTGTGGTCGCCCCGCCTGCACCTCCCTGTGCA  4343

Qy   81 Leu---------ArgThrProLeuArgGlyLeuLysPro------------  91
           :::||||||||         |||      :::||||||
Db 4342 CTGGCTGCTCCTTGGCAAACACCACCTTGGCTTCGCCACGTCCAGGCTGAAGGAGAGCC  4283

Qy   92 ThrGlyProArgSer--------ThrMetGluCysProProAlaLeuIleValHisPro  108
           |||||||:::         |||        :::
Db 4282 GCTGGCCCCCGGCCTCGCAGCTATACTCCCCGGTGTCCGCCTGGCATGCCCTGCTGCACCA  4223

Qy  109 ProAlaGlyGlyMetAlaSerGlyInPro---------TrpAla  122
           |||||||         :::             |||:::
Db 4222 CCAGCC----------TCCGTGCAGCCCACAGCTCCATTCGAGCTGGAGC  4172

Qy  123 AlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThrArgGly  142
           :::||||||:::|||    |||
Db 4171 TCAGCTTCTTCCCGTCTTGTACCACGTCCACCTCTGTCTGGGCCTGGCACCTGCAGC  4112

Qy  143 ----------------------------ProProGlnProLeu-----------------Met  149
                                          :::||||||            :::
Db 4111 TCAGCGTGGCAATGGCCCCCGGCCTGCACCTTCCTGTGTGCCAGCTGCTCCTTTG  4052

Qy  150 ArgThrPro----------------AlaAlaArgSerHisTrpProIlePro  161
           :::|||           |||:::
Db 4051 CAAACACCGGCCTTGGGCTGGCTGAAACATCCAGGTGGAAGGAGAGCCGTGGCCCCCCAGCCT  3992

Qy  162 HisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArgSer  181
           |||               :::         |||:::
Db 3991 CACAGGTGT-----------ACTCCCAGCATCTGCTGGCCTGCCTGCCTGCCACCAGCT  3941

Qy  182 ThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAla----------  198
           ||||||       |||:::             |||
Db 3940 GCCGCATGCAGCCCGCAGCCTCATGCGCACTTTCGAACTGGAGCTCAGCTTCTTCCCAT  3881

Qy  199 ProCysArgAlaGluLysLeu  205
           |||||
Db 3880 CCTTGTACCACGTCACCTCTG  3860
```

```
Qy  114 AlaSerGlySerSerGlnProTrpAlaAlaAlaSerAlaThr-------------  127
           ||||||:::          ||||||:::|||
Db  941 GCGGCTGCGGCGCCGAGCTCCAGCAACGGCGCCGTCATCGACCGCGCGGCCCAGCCCCACGG  1000

Qy  128 -------ProMetLeuSerSerLysAlaSerLeuCysIlePro-------Thr  140
                  |||                            |||     :::
Db 1001 GGAGCACCGACATGAAAGATACTGCAAGC---------CCGGGGACATGGAGGAGGATCA  1051

Qy  141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArg---------------  155
           |||||:::       ||    :::||||||
Db 1052 CGGTCTCCCGGCCGGCGATCGTGGCGTCTCGCCCGCCCAGCCGTCCCGGCAACTCGCGT  1111

Qy  155 -------------------------  155
Db
```

```
Db 1112 TCCTCAACCGCGGCGGGGTGCAGTGCCAGCCGCCGCGGCGCACGGCCGGCCTGCGCTGCACGGTCC  1171

Qy  156 SerHisTrpProIleProHisProCysAspThrAlaCysProAlaProLeuPro-------  173
           ||||||||||:::|||
Db 1172 AGGAGGTGGAGACGTCAAGATCGTGCTCATGGTGCTGGTCTTCTTGTCCACCATCA  1231

Qy  174 ValValLeuValAlaProArgSerThrIleLeuSerMetSerArg-------ThrTrp  190
           :::          ||||||:::|||||||
Db 1232 TGCTCAACTGCTGCTGGCGCAGCTGTCCACCTTCTCCGTCGAGCAAGCGGCCACCATGG  1291

Qy  191 ThrCysArgArgTrpAlaValAlaProCysArgAlaGluLysLeuMetCysSerSerSer  210
           ||||||||:::|||             :::
Db 1292 ACA--CGCGGCGTGGGAGCCTCACCGGTGCCGCCGGCCGTGCTGCCGTCTTCCCTGTCA  1348

Qy  211 ArgSer  212
           ||||||
Db 1349 CGTTCA  1354
```

```
Alignment Scores:
Pred. No.:                    56.5          Length:        23907
Score:                       162.00         Matches:          74
Percent Similarity:           34.5%         Conservative:     18
Best Local Similarity:        27.7%         Mismatches:      101
Query Match:                  14.3%         Indels:           74
DB:                             12          Gaps:             14
```

```
US-09-989-890-238 (1-212) x US-11-186-731-6 (1-23907)

Qy    2 ProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln  21
           ||||||||||                 ||
Db 4627 CCCTGCTCGCTGCTGCTTCTTGGCAAACACCGCCTTGGGCTTGGGCTTGGCCACGTGCAGGTGGA  4568

Qy   22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerPro------ProPro  39
           |||                                  ||||||
Db 4567 AGGAGAGCCGCTGGCCTCCCCGGCCGTCCCCGCGGTCTGTACTCCCCGGCCGTCCCGGCCGCCT  4508
```

```
;  LOCATION: (1)...(71)
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (72)...(23978)
;  FEATURE:
;  NAME/KEY: 3'UTR
;  LOCATION: (23979)...(24120)
US-11-186-731-4
```

```
Db            3951 CCTTGTACCACGTCACCTCTG 3931
```

```
Alignment Scores:
Pred. No.:                56.9        Length:          24120
Score:                   162.00       Matches:            74
Percent Similarity:       34.5%       Conservative:       18
Best Local Similarity:    27.7%       Mismatches:        101
Query Match:              14.3%       Indels:             74
DB:                                   Gaps:               12
```

```
RESULT 7
US-10-858-730-162
; Sequence 162, Application US/10858730
; Publication No. US2005025556A1
; GENERAL INFORMATION:
;  APPLICANT: Bailey, Richard B.
;  APPLICANT: Blomquist, Paul
;  APPLICANT: Doten, Reed
;  APPLICANT: Driggers, Edward M.
;  APPLICANT: Madden, Kevin T.
;  APPLICANT: O'Leary, Jessica
;  APPLICANT: O'Toole, George
;  APPLICANT: Trueheart, Joshua
;  APPLICANT: Walbridge, Michael J.
;  APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 162
;  LENGTH: 1590
;  TYPE: DNA
;  ORGANISM: Streptomyces coelicolor
US-10-858-730-162
```

```
Alignment Scores:
Pred. No.:                14.8        Length:          1590
Score:                   158.00       Matches:            81
Percent Similarity:       31.1%       Conservative:       18
Best Local Similarity:    25.5%       Mismatches:         89
Query Match:              13.9%       Indels:            130
DB:                                   Gaps:               15
```

US-09-989-890-238 (1-212) x US-11-186-731-4 (1-24120)

```
Qy     2 ProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGlyGln 21
           |||||||
Db  4698 CCCTGCTCGCTCGGTGCTCCTTGGCAAACACCGGCTTGGGCTTCTGCCACGTGCAGGTGGA 4639

Qy    22 ArgAlaLeuProSerSerLeuAlaLeuAlaLeuLeuSerArgProLeuSerPro------ProPro 39
           |||                                     |||      ||||||
Db  4638 AGGAGAGCCGCTGGCTCCCCGCCTCGCAGCTGTACTCCCCGGGTCCGCCTGGCCGCCT 4579

Qy    40 AlaAla--------------CysSerGlyAspProGlyCysGlySerGlyAlaGlyLeu 54
           |||                |                ||||
Db  4578 GCTGCACCACCAGCCTCCGTGTGCGACCCCACAGCCTCCATGCGCACTT-------- 4531

Qy    55 ProSerAlaSerAlaAlaAlaGlyIleAlaSerAlaValGluProValCysGly--- 73
           |||||     |:||:|||
Db  4530 ---TCGAGCTGAAGCTCAGCTTCTTCCCGTCCTTGTACCACATCACCTCCGTCTGGGCCT 4474

Qy    74 ------------------------------AspAlaAlaProAlaCysLeu 80
                                          ||||||
Db  4473 GAGCCACCTCGCAGCTGAGTGTGTGCTGCGCCCCGCCTGGGCCTGCACCTCCCTGTGCA 4414

Qy    81 Leu--------ArgThrProLeuArgGlyLeuLeuLysPro----- 91
           |||        ||                          ::|||
Db  4413 CTGCTGCTGCCTCCTTGGCAAACACCACCCTTGGGCTCTGCCCAGGCTGAAGGAGAGGC 4354

Qy    92 ThrGlyProArgSer--------------ThrMetGluCysProProAlaLeuIleValHisPro 108
           |||                         |||            :::
Db  4353 GCTGGCCCCCGGCCTCGCAGCTATACTCCCCGGTGTCCGCCTGCATGCCGTGCTGCACCA 4294

Qy   109 ProAlaGlyGlyMetAlaSerGlySerGlnPro------------TrpAla 122
           ||                         :::           |||:::
Db  4293 CCAGCC--------TCCGTGTGCGACCCCACAGCCTCCATTCGAACTTTCGAGCTGGAGC 4243

Qy   123 AlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThrArgGly 142
           :::|||||:|||                 |||             |||
Db  4242 TCAGCTTCTTCCCGTCCTTGTACCACGTCACCTCTGTCTGGGCCTCACCTCGCAGC 4183

Qy   143 ------------------ProProGlnProLeu-------- 149
                             ||||||                        |||
Db  4182 TCAGCGTGGCAATGGCCCCCGGCCTCGCACCTTCCTGTGCCAGCTGCTCCTTTG 4123

Qy   150 ArgThrPro------------------AlaAlaArgSerHisTrpProIlePro 161
           :::|||||                  ||||||
Db  4122 CAAACACCCGCCTTGGCTCTGAAACATCCAGTGGAAGGAGAGCCGCTGGCCCCCAGCCT 4063

Qy   162 HisProCysLysAspThrAlaProLeuProValLeuValAlaProArgSer 181
           |||                |||      :::           |||:::
Db  4062 CACAGGTGT--------ACTCCCCAGCATCTGCCTGCCTGCTGCTCACCACCAGCT 4012

Qy   182 ThrIleLeuSerMetSerArgThrTrpThrTrpCysArgArgTrpProAlaValAla-------- 198
           |||:::                 |||::|
Db  4011 GCCGCATGCAGCCCGCCAGCCTCTATGCGCACTTCGAACTGGGAGCTCAGCTTCTTCCCAT 3952

Qy   199 ProCysArgArgGluLysGlyLeu 205
           |||||
```

US-09-989-890-238 (1-212) x US-10-858-730-162 (1-1590)

```
Qy    12 ThrProArgThrLeuAlaThrMetGlyGlnArgAlaLeuProSerSerLeuAlaLeuLeu 31
           |||||||                                ||:::||
Db   281 ACGCCCCGACCT--------------------------CCAACATCGTCACCGCGCCG 313

Qy    32 SerArgProLeu-------SerProProAlaAlaCysSerGlyAspPro------- 46
           |||                                      |||
Db   314 AGCTGGCCCTGCGGCCTGCTGCGTGCGCACCGCACCGCCCGCAACATCCCGAGGCCAACGCCGCGGC 373

Qy    47 --------GlyCysGlySerGlyAlaGlyLeu--------------------Pro 55
                  |||                                              |||
Db   374 TGAAGAACGGCGAGTGGAAGCGCAGCAAGTACACCGGCGTGGCCGGAGAGACCC 433

Qy    56 SerAlaSerAlaAlaAlaGlyIleAlaSerAlaValGluProValCysGlyAspAla 75
           |||||||:::                      :::|||            :::
Db   434 TCGGCGTCGTCGGGCCTCGGCCCATCGGGCCAGCCGCATGTCGGGCCTTCG 493

Qy    76 Ala----------------ProAlaCysLeuLeuArgThrProLeuArgGlyLeu 88
           |||                 |||
Db   494 GCATGAAGGTCGTCGCCTACGACCCCCTACGTGCAGCCGCGGCGGCCGATGGGCG 553

Qy    89 LeuLys---------------------------ProThrGlyProArgSerThrMet 98
           |||||                                        |||
Db   554 TCAAGGTGCTGTCCCTGGACGAGCTGCTGGAGGTCTCCGACTTCATCACGGTCCACCTGC 613

Qy    99 GluCysProProAlaLeuIleValHisProProAla------------ 110
           |||:::                    |||
Db   614 CCAAGACCCCCG------------------------AGACCCTCGGCCTGATCGGGCAGCAGGCGCGTCGCA 661

Qy   111 ---GlyGlyMetAlaSerGlySerSerGlnProTrpAlaAlaAlaSerAlaThrProMet 129
           |||
```

DB:

US-09-989-890-238 (1-212) x US-10-858-730-36 (1-2625)

```
                                                              8                              Gaps:            13

Qy      1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
           ||||||       :::    :::   |||
Db   1394  AGTCCGCTGACGACATCGCCGCCGTCTACGAGCTCGACGAGCACGCACCCCGGGCA  1453

Qy     21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProPro---  39
           ||||||                      -------Thr  140
Db   1454  AGGCGC------------------CCATCCTCGACGTCA------TCCCGCTCTTCGAAACCGGTGCCGACC  1501

Qy     40  ---------------------AlaAlaCysSerGlyAlaAspProGlyCysGlySerGly  51
                                ||||||
Db   1502  TGGACGGCGGCCCCCAGGTCCTGGGCGGCGTCCCCGGCCGTCCAGCGCGCGCC  1561

Qy     52  AlaGlyLeuProSerAlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluProVal  71
           :::      |||||::::|||         :::    :::|||||
Db   1562  TCGAGCAGACGGCCGCGCATGGAGTCATGCTCGGCTACAGCGACTCGCCAAGGACG  1621

Qy     72  CysGlyAspThrAlaCysProAlaProAlaCysTyrLeuLeuArgGlyLeuLeuLysPro  91
           |||          ||| :::|||                         ------CCG  1675
Db   1622  TCGGCCCGGTCAGGCGACCTGCGGGCTCTCACACCGCCCCCAGGCGCGGCTGG------  1675

Qy     92  ThrGlyProArgSerThrMetGluCysProPro-------------AlaLeuIleValHis  107
                                                         ||:||
Db   1676  AATGGGCGCGGAGCACGACGAGCATCAAACTCACCCTGTTCCACGGCGCGGTGCCCTGG  1735

Qy    108  ProProAlaGlyGlyMetAlaSerGlySerSerGlnPro-------------TrpAla  122
           |||                                                ||||:::
Db   1736  GCCGCGGCGGCGGGCCCCCAACCGGCGCGTCCTCGCCCAGGCCCCGGATCGGTGGACG  1795

Qy    123  AlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThrArgGly  142
           :::    :::|||||||
Db   1796  GCCGGTTCAAGGTCACCGAGCAGGGCGAAGTCATCTCG----------------------  1834

Qy    143  ProProProGlnProLeuMetArgThrProAlaAlaArgSerHis---------------  157
           |||:::|||||:::                               ---------------
Db   1835  ---CCCGCTACGGTCAGGCGGCGGATCGCCCACATCGAACAGGTGGGCCACGCG  1891

Qy    158  ----------TrpProIleProHisProCysAsp-------------------------  165
                     ||||||
Db   1892  TGCTCATGGCCTCCACCGAAAGCGTGCAGCGGAGCGGAGAGCCCGCGCCCGTTCC  1951

Qy    166  ----------ThrAlaCysProAlaLeuProValValLeuValAlaProArgSer  181
                     ||||||
Db   1952  GCGGTATGCGCATCGCCGAAGCCG-----CCCACGCGCGCCCTACCGCGCCCTCGTCG  2008

Qy    182  ThrIleLeu---SerMetSerArgThrTrpThrCysArgArgTrp----AlaValAlaPro  199
           |||           :::                             :::
Db   2009  ACACTGAAGGGTTCGCGGAGTGGTTCTCCCGGGTGGAGGAGAGCTGAGTGAGC  2068

Qy    200  Cys---------------------ArgAlaGluLysLeuMetCysSerSerArgSer  212
           |||                    ||||||                ::::|||
Db   2069  TGCGGCGTGGGTCGGCTCCGGCGGTCCGGCGGCCTCGACGACCTCC  2125
```

RESULT 8
US-10-858-730-36

```
Alignment Scores:
Pred. No.:                23.1
Score:                   157.00
Percent Similarity:       37.5%
Best Local Similarity:    27.8%
Query Match:              13.8%

Length:              2625
Matches:               72
Conservative:          25
Mismatches:           100
Indels:                62
```

RESULT 9
US-11-136-527-3025

DB:

```
Db    662  AGGTCAAGCGAGCGGTCCGCATCGTCAACGCCCGCGGCGGCATCGTCGACGAGGAGG  721
           |||||    |||     |||    |||||||||||:::|||

Qy    130  LeuSerSerLys-----------------------------AlaSerLeuCysIlePro  139
                ||||||::                         |||||
Db    722  CGCTGTACTCGGCGCTCAAGGAGGGCCGCCGTGCCGGCGCCGCCTCGACGTGTACGCCA  781

Qy    140  --------------------------------------------------Thr  140
                                                             |||
Db    782  AGGAGACCCTGACCACGGACCTCGCCGTGTTCGAGTTCGACCAGGTGGTGCCACCCCGACC  841

Qy    141  ArgGlyProProGlnProLeuMetArgThrProAla----------------AlaArgSer  156
           ||||||||||||                                       :::|||
Db    842  TCGGCGGCCTTCCACCGACGAGCCCCAGGAGGCCGGCATCGCCGGTCGCCAAGTCGGTCC  901

Qy    157  HisTrpPro----------------------------------------------Ile  160
           |||                                                   |||
Db    902  GCCTGGCCCCTCGCCGGTGAGCTGGTCCCCGACGCGGTCAACGTCGCAGGCGGTGTCATCG  961

Qy    161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg  180
           |||          :::||| :::|||
Db    962  CCGAGGACGTCAAGCCCCGGTCTCGCCGCTCGCCGAGCGCCTCGGCCGCATCTTCACCCGCGC  1021

Qy    181  SerThrIleLeuSerMetSer---------------------------------------  187
           ||||||
Db   1022  TCGCGGGGTGAGGTCGCCCGTCCGGCCTCCGACGTCGGAGGTCGAGGTCTACGGCGAGATCACCCAGCACG  1081

Qy    188  -----ArgThrThrCys-----------------------ArgArgTrpAlaValAlaAla  198
                ||:::                              :::
Db   1082  ACGTGAGGTGCTGGAGCTGTCCGCCCTCAAGGACGTGCTTCGAGGACGTCGACGAGA  1141

Qy    199  ProCysArgAlaGluLysLeuMetCysSer-------------------SerSerArgSer  212
           ||||||                                           ::|||||||
Db   1142  CGGTGTCGTACGTCAACGCCCCGTGTTCGCCCAGGAGCGCGGCGTCGAGGTCC  1195
```

```
;  TYPE: DNA
;  ORGANISM: Rattus norvegicus
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (690)..(690)
;  OTHER INFORMATION: n is a, c, g, or t
US-11-136-527-3025

Alignment Scores:
Pred. No.:                        3334
Score:                            157.00
Percent Similarity:    36.2%      Matches:          71
Best Local Similarity: 30.6%      Conservative:     13
Query Match:           13.8%      Mismatches:       89
DB:                               Indels:           59
                                  Gaps:             11

US-09-989-890-238 (1-212) x US-11-136-527-3025 (1-3334)

Qy    22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerArgProProProAlaAla 41
            |||||||||||||::::|||||||||||
Db   810 CGAGGGCCTTCCCAGCGCCTTCCGGCTCTCCTGCGCCCGGGCTCTCCTGAGCCCCGGCGG 869

Qy    42 CysSerGlyAspProGlyCys------GlySerGlyAlaGlyLeuProSerAlaSerAla 59
            :::|||||         |||||||    |::: :::
Db   870 GCCACCGGCCTGCTTGGCTGCGGGTTCGGGTCTGGCGTGGGATGTTAGCTGTGGGGGCG 929

Qy    60 AlaAlaGlyIleAlaSerSerAlaVal----------GluProValCysGlyGlyAspAlaAla 76
            ||||        |||
Db   930 ATGGAGGGCCCCGCCAGAGCCGTTCCTGCTCAGCAGCCCGCCCTTGGCCGCCCTGCAC 989

Qy    77 ProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeu-------LysProThrGly 93
            :::::|||||||||
Db   990 AGCATGGCCGAGATGAAGACCCCGTCTACCCTGCCGCCTATCCCCGCTGCCACCGGGG 1049

Qy    94 ProArgSerThrMetGluCys--------------ProProAlaLeuIleVal 106
            |||:|||
Db  1050 CCCCCCTCTCCTGTCCTTCCTCGTCCTCGTCCTCGCCTTTGGGGCCA 1109

Qy   107 His----------------ProProAlaGlyGlyMetAlaSer---GlySer--- 117
            ||||           |||::::|||
Db  1110 CACAACCCGGCGGTTGAAGCCCCGGCCGGGGGGCCTCTCGTCCCTGGGCAGCCCC 1169

Qy   118 -------------------------------------Ser 118
Db  1170 CCCGCAGCAGCTTTCGGCGGCCACTCCGCACGGCATCAACGACACATCCTGAGCCGGCCCTCC 1229

Qy   119 GlnProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSer-LeuCysIl 138
            |||:::   :::||||::   |||
Db  1230 ATGCCGGTGCGGTCTCGGGGACCGGCCGCCCTCCGGGGTCTTCTTCCTCC 1289

Qy   138 eProThrArgGlyProProGlnProLeuProMetArgThrProAlaAlaArgSerHisTr 158
            |||:::|||
Db  1290 TCCTCTCCGTCCGCCTCCGCCACTTCGGCCTTCGCGGCCGCGTGCTG-------CT 1340

Qy   158 pProIleProHisPro-----------------CysAspThrAlaCysProAl 170
            |||::|||||   |||
Db  1341 GCCCTCCGCGCGCCGCTCCGCCCTCATCGCCCGGCGTGTTGGCCCGGCCTGCCCGC 1400

Qy   170 aProLeuProValVal-----ValAlaProSerThrIleLeuSerMetSerAr 188
            |||||||||
Db  1401 TTCAGCAGCCTGAGCCCTCCGCCACCGCCCCGGGCTCTACTTTAGCCCCAGCCCCGCG 1460

Qy   188 gThrTrpThrCysArgArgTrpAlaValAlaPro 199
            |||||||
Db  1461 GCCGTGGCCG------CCGTGGGCCGGTACCCCA 1488

RESULT 10
US-10-330-773-278/c
; Sequence 278, Application US/10330773
; Publication No. US2006040262A1
; GENERAL INFORMATION:
;   APPLICANT: David W. Morris
;   APPLICANT: Marc Malandro
```

```
;  TITLE OF INVENTION: Novel Compositions and Methods in Cancer
;  FILE REFERENCE: 52945001300
;  CURRENT APPLICATION NUMBER: US/10/330,773
;  CURRENT FILING DATE: 2002-12-27
;  NUMBER OF SEQ ID NOS: 981
;  SOFTWARE: FastsEQ for Windows Version 4.0
;  SEQ ID NO 278
;  LENGTH: 127722
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(127722)
;  OTHER INFORMATION: n = A,T,C or G
US-10-330-773-278

Alignment Scores:
Pred. No.:                        127722
Score:                 156.50      Length:           127722
Percent Similarity:    40.1%       Matches:          53
Best Local Similarity: 32.7%       Conservative:     12
Query Match:           13.8%       Mismatches:       54
DB:                                Indels:           43
                                   Gaps:             6

US-09-989-890-238 (1-212) x US-10-330-773-278 (1-127722)

Qy    22 ArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerPro----------- 37
            ||||||||||||||||:::: |||    ||| ||||::::
Db 10147 CGGGCACTGCCAGCCGGCGGCGGCTCCCGGCTCCCGCCGCTCCCGGCCTCCGGGC 10088

Qy    38 --------ProProAlaCys------------------SerGlyAsp 45
                    |||
Db 10087 CGGCCTTCCGGGCCGCAGCAGCGGCGGCTCATTACCGTATGGCCGGTGGGGTCGGGGCCGG 10028

Qy    46 ProGlyCysGly-SerGlyAlaGlyLeuProSerAlaSerAlaAlaAlaGlyIleAlaSe 65
            |||      :::::   :::::|||||||||
Db 10027 CCCGGCTGCGGCCCTGGTGCCGGCGCCGTCCCGGCTCCGTCCTCGTCTCGCCGCGC 9968

Qy    65 rSerAlaValGlu---ProValCysGlyAlaAspAlaAlaProAlaCysLeuLeuArgThrPr 84
            :::::|||                |||
Db  9967 CGCCGCCACTGCCAGCCCCTCCTCCCGCCGCCTGCCCTGCCTCGCTCGCTCG-- 9910

Qy    84 oLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCysProProAla-L 104
            ||||||
Db  9909 -----------------------------------------CTCCCCCCTCCGC 9896

Qy   104 euIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnProTrpAlaAlaA 124
            ||           ||:::::|||
Db  9895 TCGCTGTGCGGCGTCCCTCTCCGCCCTCCGCCTCTCCGCCGGGGTCCCCCCTCCC---- 9846

Qy   124 laSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProP 144
            |||||||
Db  9845 ---------ACCCCTTCTTCCTCCTCCTTCCCTCCTCCGGCCGCCGCGGGGCCCA 9794

Qy   144 roProGlnProLeuProMetArgThrProAlaAlaArgSerHisTrpProIleProHisProC 164
            |||||||                       :::
Db  9793 GGCCTTCCGCGCGCCGCTCCGCCCCCCCG------CGGCCGCCGCGGCCCCCGGCGCT 9740

Qy   164 ys 164
            ||
Db  9739 GC 9738

RESULT 11
US-11-128-061-745/c
; Sequence 745, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
;   APPLICANT: Melville, Mark W.
;   APPLICANT: Charlebois, Timothy S.
;   APPLICANT: Mounts, William M.
;   APPLICANT: Hann, Louane E.
;   APPLICANT: Sinacore, Martin S.
;   APPLICANT: Leonard, Mark W.
```

```
Db  6352  TCAGCGTCGTCAGCTCGTCCACCTGGGCCTTCTGCCGACGCGTCTCGTCTCGCAAACT 6293
               |||||          ||  ||||||  |||||
Qy    71  ValCysGlyAspAlaAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLys  90

Db  6292  TCTTGTGCTTCTCCATCTCCGGCGTCGGCTGCTTGCTTCT---------------  6254
               :::            |||
Qy    91  ProThrGlyProArgSerThrMetGluCysProAlaLeuIleValHisProProAla 110
               ||||||
Db  6253  ----------------GCTTCAACGCTGCTGCTTCCGCCTGTCCGACGCCGCCTCCTGCT 6206
                               |||||||||
Qy   111  ----------------GlyGlyMetAlaSerGlySerSer 118
                                 |||  :::
Db  6205  CAGCTTCCTACGCAGCTTCTCTGCAGCCGACCTGGGCCCGGGCCCTGGGGCCTGCT 6146
               :::
Qy   119  GlnProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIle 138
               :::
Db  6145  CCTCCGCCGACTGTTTCAGCGCGCTTCCGCCTCTT----------CCACTTGTTTCC 6101
                   |||
Qy   139  ProThrArgGlyProProGlnProLeu----------MetArgThrProAlaAlaArg 155
                   :::|||
Db  6100  GAGACTGAGCCGCCTTCCGTTCGGCGCCTGCTCCCGGCCTCCTCCGCCTCCGGCAGCTC 6041
               |||
Qy   156  SerHisTrpProIleProHisProCys---------------AspThr 166
               :::  ||||||||||
Db  6040  GCCGGCGCCGCCTCTGCCTCCACCCCGCAGCCGCTCCAGCATGNNNNNNNNNNNNNNN 5981
               :::
Qy   167  AlaCysProAlaProLeuProValValLeuVal----------AlaProArgSerThrIle 183
               :::
Db  5980  TCTGCTGCAGCTCCTCTTCCCGCTGCTGCACCACAAAGGCATGCGGCCTTC-TCCTCT--- 5925
               ||||||||||
Qy   184  LeuSerMetSerArgThrThrTrpThrCysArgArgTrpAlaValAlaProCysArgArgGlu 203
               :::   ||||||||||
Db  5924  --GCCTGCAGTCGCTTCTGGCAGCC----TCCTGGGCCAGCTGCAGCTGCCTGCGCGGAC 5871
                   ||||||||||
Qy   204  LysLeuMetCysSerSerArgSer 212
               ||||:||||||||
Db  5870  TC-CTGCTCGTCGTCTCGTTCCTCCGCAGGCG 5845
```

RESULT 12
US-11-128-049-745/c

Alignment Scores:
```
Pred. No.:                              Length:              14619
Score:                   85.8           Matches:                66
Score:                 155.50           Conservative:           25
Percent Similarity:     39.7%           Mismatches:             82
Best Local Similarity:  28.8%           Indels:                 58
Query Match:            13.7%           Gaps:                   10
DB:                       12
```

US-09-989-890-238 (1-212) x US-11-128-061-745 (1-14619)

```
Qy    21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerPro---ProPro 39
               ::|||  |||          ||||||
Db  6472  AGAAGAGCTCCTCTACCTGGCTACGCTGCCGGGCAGCCTCGGTCACCTCGGCCTTCA 6413

Qy    40  AlaAlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla 59
               |||||      :::  ||||||        |||:::|||
Db  6412  GCCGCTGCAGCTCCTGCTCCAGGATGCTTCTTCTGGTGGTCAGTCTCCTCCAGCTGCAGCC 6353

Qy    60  AlaAla---------------GlyIleAlaSerSerAlaValGluPro 70
               ::|||                   |||:::|||:::::
Db  5870  ...
```

```
;  NAME/KEY: misc_feature
;  LOCATION: (6573)..(6594)
;  OTHER INFORMATION: n is a, c, g, or t
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (7283)..(7317)
;  OTHER INFORMATION: n is a, c, g, or t
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (7320)..(7334)
;  OTHER INFORMATION: n is a, c, g, or t
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (9420)..(9510)
;  OTHER INFORMATION: n is a, c, g, or t
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (12124)..(12151)
;  OTHER INFORMATION: n is a, c, g, or t
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (13563)..(13585)
;  OTHER INFORMATION: n is a, c, g, or t
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (14598)..(14619)
;  OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-745

Alignment Scores:
Score:                      85.8       Length:        14619
Percent Similarity:         155.50     Matches:       66
Best Local Similarity:      39.7%      Conservative:  25
Query Match:                28.8%      Mismatches:    82
DB:                         13.7%      Indels:        58
                            12                Gaps:     10

US-09-989-890-238 (1-212) x US-11-128-049-745 (1-14619)

Qy     21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerPro---ProPro    39
               |||||||      :|:||    ||||     ||||||   :::  ||||
Db   6472  AGAGAGCTTCCTCCTTCTACCTGGCTACGGTGCCGGGACAGCTCGGTCACCTCGGCCTTCA   6413

Qy     40  AlaAlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla    59
               |||||||||||||||||||||    |||||||||    |||||||||||||||
Db   6412  GCCGCTGCAGCTCCGTCCGTCCAGGATGCTCTTCTGGTGCGTCAGTCTCCTCCAGCTGCAGCC   6353

Qy     60  AlaAla--------------------GlyIleAlaSerSerAlaValGluPro    70
               ::|||                  ||:::||||||||||::::::
Db   6352  TCAGCGTCGTCAGCTCCTCAGCTCCTCCACCTGGGCTCCTGCTCTGCTCTGCAAACT   6293

Qy     71  ValCysGlyAspAlaAlaProAlaAlaCysSerLeuLeuArgThrProLeuArgGlyLeuLeuLys    90
               ||||||||||||||||      ::|:||||||
Db   6292  TCTTGTGCTTCTCCATCTCCGGTCGGCTGGCTTGCTTCT-----------------   6254

Qy     91  ProThrArgGlyProArgSerThrMetGluCysProProAlaLeuIleValHisProProAla    110
               ||||||:      ||||:||||:|||||:||||||:||:||
Db   6253  --------------GCTTCAACGTCCTGCTCCGCCTGTGCCCACGCGCTGCCTCCTGCT    6206

Qy    111  -------------------GlyGlyMetAlaSerGlySerSer    118
                                   |||      :::
Db   6205  CAGCTTCCTTACGCAGCTTCTTCTGCAGCCGGTCCGGGGCCTGGGCCTGGCCTGGCCTGCT    6146

Qy    119  GlnProTrpArgAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIle    138
               |||||:      :::::|||:      ||||||
Db   6145  CCTCCGCCGACTGTTCAGCCGCTCGCCCTCTT---------CCACTTGTTTCC    6101

Qy    139  ProThrArgGlyProProGlnProLeu--------MetArgThrProAlaAlaArg    155
               ||||||:      ||||:::
Db   6100  GAGACTGAGCGCCGCTCCCCGTTCGGCTCGCTGCCTCCTGCTCCCGGGCCTCCTCGGCAGCTC    6041

Qy    156  SerHisTrpProIleProHisProCys--------------------AspThr    166
               :::    ::|:||||||
Db   ...
```

```
Db   6040  GCCGCGCCGCCTTCGCTCACCCGAGCCGCTCCAGCATGNNNNNNNNNNNNN   5981
               |||||||||||||||||
Qy    167  AlaCysProAlaProLeuProValValLeuVal--------AlaProArgSerThrIle   183
               :::                                |||||
Db   5980  TCTGCTGCAGTTCCTCTTCCCGTCTGCCACCACAAAGGCATGCGCCTTC-TCCTCT-   5925

Qy    184  LeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCysArgAlaGlu   203
               :::
Db   5924  ---GCCTGCAGTCGCTTCTGGGCAGCC---TCCTGGGCCAGCTGCAGCTGCCGGAC   5871

Qy    204  LysLeuMetCysSerSerSerArgSer   212
               |||||:|||
Db   5870  TC-CTGCTCTGCTCGTTCTCGCAGGCG   5845

RESULT 13
US-11-096-568A-23836/c
;  Sequence 23836, Application US/11096568A
;  Publication No. US20060048240A1
;  GENERAL INFORMATION:
;  APPLICANT: Alexandrov, Nickolai et al.
;  TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
;  TITLE OF INVENTION: Therby
;  FILE REFERENCE: 2750-1592PUS2
;  CURRENT APPLICATION NUMBER: US/11/096,568A
;  CURRENT FILING DATE: 2005-04-01
;  NUMBER OF SEQ ID NOS: 34471
;  SEQ ID NO 23836
;  LENGTH: 1835
;  TYPE: DNA
;  ORGANISM: Zea mays subsp. mays
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(1835)
;  OTHER INFORMATION: Ceres Seq. ID no. 1415689
US-11-096-568A-23836

Alignment Scores:
Pred. No.:                  24.3       Length:        1835
Score:                      154.50     Matches:       65
Percent Similarity:         35.7%      Conservative:  15
Best Local Similarity:      29.0%      Mismatches:    79
Query Match:                13.6%      Indels:        65
DB:                         9                Gaps:     10

US-11-096-568A-23836 (1-1835) x US-11-096-568A-23836 (1-1835)

Qy      8  ProValAspGlnThrProArgThrLeuAlaThrMetGlyGlnArgAlaLeuProSerSer   27
               |||                                    :::
Db   1600  CCTGGTCCCGCCGACATGCCGGCCGGCCAGCTGCCACTGGAACGCGGTGCA---CCAGCGACG   1544

Qy     28  LeuAlaLeuLeuSerArgProLeuSerPro----------------------    38
               |||
Db   1543  CTA------------GAACCAGCGACGCACCCGCTCCGCCGCCTCCGCCTGGGCACA   1496

Qy     39  ----------------------ProAlaAlaCysSerGlyAspProGlyCys    48
                                      :::
Db   1495  GCCTACGGGCCGGAGCCGAGCCGAACGGGACGAACTTCAGCTGCTTCCCGGGAAGTTCAGGTCCC   1436

Qy     49  GlySerGlyAlaGlyLeuProSerAlaSerAlaAlaAlaAlaGlyIleAlaSerSerAlaVal   68
               |||                      ::::
Db   1435  TGGCCAGGA----------------ACCGCTCCGGCAGGAACTCGTCGGGCCTCTCCC   1394

Qy     69  GluProValCysGlyAspAlaAlaProAlaAlaCysLysLeuLeuArgThrProLeuGlyLeu   88
               |||
Db   1393  ACGCCGCCGGGGTCGGCATGATCGCC---------ACGAGTTGAAGATCACGG   1349

Qy     89  LeuLysProThrArgGlyProArgSerThrMetGluCysProProAlaLeuIleValHisPro   108
               |||                           :::
Db   1348  TGCAACCCCTGGGCACCGGCCTAGCCGCTATCTGTACCC------------CGTCCT   1304

Qy    109  ProAlaGlyGlyMetAlaSerGlySerGlnProTrpAlaAlaAla----------Ser   125
               |||      :::         :::
Db   1303  CCACGGGCCTTGTGCGGCAGCAACACGGGCCCGGGTGCAGCCCATGGCCATGGCCTCCCTCA   1244
```

```
Qy   126 AlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThrArgGlyProProPro 145
             |||                              |||||||
Db  1243 CGACGGCCT----------------------------GCAGGACGGCAGCCCGGCCCGCGT 1211

Qy   146 GlnProLeuMetArgThrProAla---------------------AlaArgSerHisTrp 158
             :::|||                                       |||
Db  1210 CGGCCTCCTCGACGGCGTCCTGCGCCCAGCGCGTCCCTGACCTCCCGCCAGCCTGG 1151

Qy   159 ProIleProHisProCysAspThrAlaCysArgArgTrpAlaValValLeuValAla 178
             ||||||||
Db  1150 CCATCACGCCCGGGTTGCGGAGCAGCTCGGCCATCGCCC--------------ACA 1109

Qy   179 ProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAla 198
             |||||                       :::              :::
Db  1108 CCACGGTCAGCGGCCATGGTGTCGCTCCCCGGCCGGAAGATGTCGAAGAGACGGTCAGCGTCA 1049

Qy   199 ProCysArgAla 202
             |||||
Db  1048 CGTTGTCGCGCG 1037
```

RESULT 14
US-11-096-568A-21912
; Sequence 21912, Application US/11096568A
; Publication No. US2006004824OA1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 21912
; LENGTH: 1269
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1269)
; OTHER INFORMATION: Ceres Seq. ID no. 12406782
US-11-096-568A-21912

```
Alignment Scores:
Pred. No.:               20.1          Length:        1269
Score:                   154.00        Matches:       67
Percent Similarity:      36.7%         Conservative:  24
Best Local Similarity:   27.0%         Mismatches:    96
Query Match:             13.6%         Indels:        61
DB:                      9             Gaps:          9
```

US-09-989-890-238 (1-212) x US-11-096-568A-21912 (1-1269)

```
Qy    12 ThrProArgThrLeuAlaThrMetGlyGlnArgAlaLeuProSerLeuAlaLeuLeu 31
             |||||||||
Db   161 ACCCCGCGGC------------------CGCCCCCTCCCCACCACCCTCGCCATGGTC 199

Qy    32 SerArgProLeuSerProProAlaAlaCysSerGlyAspPro--------------- 46
             ::|||||||
Db   200 AGCGCCGCCGCCTTCGCCCCGCCG---------TCCGGGAAGCCGACCGACGAGCTG 250

Qy    47 --------------------------------------------GlyCysGlySer 50
                                                          |||
Db   251 AAGCGCTTGGCGGGCACGGGCCGCGTGGAGCCCGTCGTGGAGCGGCTCGGCGGCCTG 310

Qy    51 GlyAlaGlyLeuProSerAlaSerAlaAlaAlaGlyIle-AlaSerSerAlaValGluPr 70
             ::|||||||:::
Db   311 GGCACGGGCCTCCACGGGCCACGCGTGACCCGGCCTTCCTCCGGCGGGC 370

Qy    70 oValCysGlyAspAlaAlaAlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLy 90
             :::|||:::
Db   371 GCGCTGC---CGGGGGTGTCCGGCGGTGCCGACCTCCCGCACCGCGGAGGCGCACGCGGCG 427
```

```
Qy    90 sProThrGlyProArgSerThrMetGluCysProProAlaLeuIleValHisProProAl 110
             :::                              |||||||
Db   428 CGGCTCGGCATCCGCTCCCCGCTGCACGCGCCGACACGCGTCATCGACACGGC 487

Qy   110 a-------------------------------------GlyGlyMetAlaSerGlySerSe 118
         |
Db   488 GCCGACGAGGTCGACCCGACCTCAACCTCGTCAAGGGCGCGGCGGCTCGCTGCTCCGC 547

Qy   118 rGlnProTrp-------------AlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLe 136
             :::        |||
Db   548 GAGAAGATGTCGAGGCGCCGGCGCCCGCCGCTTGCGTCATCGTCGACGAGTCCAAGCTC 607

Qy   136 uCysIleProThrArgGlyProProGlnInProLeuMetArgThrProLeuAlaAlaArg- 155
             |||||                    :::              :::
Db   608 GTGCCGCGCCTCGGCGTGCACCGGCCCGTCCCGTGCGAGGTCGGGCGCGCCC 667

Qy   156 -SerHisTrpProIleProHisProCysAspThrAlaCysProAla--------------- 170
             ||||||
Db   668 CACACGCTGGGCCTCATCGCCGGCCTTCCACGGCCTGCCGCGCCTTCGTCACCGAC 727

Qy   171 --------ProLeuProValValLeuValAlaProArgSerThrIleLeuSerMetSe 187
             :::
Db   728 AGGACCACCACCGTCCCCGAGGGGGAGCAAGGCTCCGAGCAGCAGCCCTTCGTCACCGAC 787

Qy   187 rArgThrTrpThrCysArgArgTrp----------AlaValAlaProCysArgAl 202
             :::
Db   788 AACGGCAACTACATCGTCGAGATGTTCTTCGACGCGCATCCGCGGCCATCCTGCGCGAC 847

Qy   202 aGluLysLeuMetCysSerSer 209
             |||||
Db   848 ATCAGCGACGCCGCCTGCTCCGCA 869
```

RESULT 15
US-11-096-568A-19912
; Sequence 19912, Application US/11096568A
; Publication No. US2006004824OA1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 19912
; LENGTH: 1676
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1676)
; OTHER INFORMATION: Ceres Seq. ID no. 12375592
US-11-096-568A-19912

```
Alignment Scores:
Pred. No.:               24.2          Length:        1676
Score:                   154.00        Matches:       63
Percent Similarity:      35.4%         Conservative:  21
Best Local Similarity:   26.6%         Mismatches:    113
Query Match:             13.6%         Indels:        40
DB:                      9             Gaps:          7
```

US-09-989-890-238 (1-212) x US-11-096-568A-19912 (1-1676)

```
Qy     1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
             |||:::|||||||
Db   522 TCCCCGTCCCGGCGCCGGCCGCTACCCGGACGACGACCACCAGCATCCTGTCTCCACCCGC 581

Qy    21 GlnArgAlaLeuProSerLeu--------------------------------------- 28
             |||
Db   582 TCCCGCCACGAGCAGCTTCCGATCGACCCCCGCCCCTGCGCCGGCCACCTCCGCTCCTGGC 641

Qy    29 -------AlaLeuLeuSerArgProLeuSerProProAlaAlaCysSerGlyAsp 45
```

```
Db                   |||         :::    |||    |||||||||                    ||||||
       642 ACCGCCTCCGCCGGCTACGACCCTGCCGTCCTCCGCGCCTCCCCGCTCTCCATCC 701

Qy    46 ProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAlaAlaGlyIleAlaSer 65
                   ||||||||||           |||||::::::     ::::
Db       702 CCGGGCGGCATCCCCTCCTCG------CCGTCGTCGGCAACAGCGGCATCCTCCTGCGCG 755

Qy    66 SerAlaValGluPro---------ValCysGlyAspAlaAlaProAlaCys 79
                   :::||||||                   :::
Db       756 CGAACCATGGAGCCCATCGATTCCCACGCAGCCGTGTTCCGCCTCAACAACGCCCGCG 815

Qy    80 LeuLeuArgThrProLeuArgGlyLeuLeuLysProGlyProArgSerThrMetGlu 99
                   ||||||      :::||||||   |||  ||||||
Db       816 TCTCAGGCTACGCCGCCCGCACGTTGGCAGCAAGACCAACTTCTCCTTCATCAACAGCAACG 875

Qy    100 CysProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGln 119
                   :::        ::   :::::::
Db       876 TTCTTCACCTCTGCGCGCCCGGCTGCTTCTGCCACCCTTACGGCCACGGCGTCC 935

Qy    120 ProTrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaAlaSerLeuCysIlePro 139
                   |||    :::    ||||||         :::
Db       936 CCATCCTGCTCTACACATCTGCCAGGCGCCGCCATTTCGTCGACGTCG---------- 980

Qy    140 ThrArgGlyProProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpPro 159
                   |||      |||
Db       981 ----------CCGCCTGCAACGCCACCTCGTCTCCCGCCACCGGCTCCCCCATCTCCG 1028

Qy    160 IleProHisProCysAspThr---------AlaCysProAlaProLeuProValVal 175
                   ||||||                 :::   |||
Db       1029 TCACCGACGCCCGCGCCTCGACGTCCTCTGCGCGGCATCGTCAGTACTACTCCCTCCGCC 1088

Qy    176 LeuValAlaProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrp 195
                   :::||||||                 :::::       |||
Db       1089 GATTCGTCGCCGAGACCGGCC------GCGCGCCCGAGGAGTGGGACCCGCGGC------ 1136

Qy    196 AlaValAlaProCysArgAlaGluLysMetCysSerSerSerArgSer 212
                   ||||||       :::    |||   |||
Db       1137 ACGACGCCGCCCATGTTCCATTACTCGTCCGGATGCAAGCAAGCCATCATGGTCG 1187
```

Search completed: March 17, 2006, 06:26:57
Job time : 1524 secs

THIS PAGE LEFT BLANK

```
OM protein - nucleic search, using frame_plus_p2n model

                GenCore version 5.1.7
        Copyright (c) 1993 - 2006 Biocceleration Ltd.

Run on:      March 17, 2006, 06:01:44 ; Search time 4842 Seconds
                                         (without alignments)
                                    2488.809 Million cell updates/sec

Title:       US-09-989-890-238
Perfect score: 212
Sequence:    1 SPHQAAAPVDQTPRTLATMG.......RRWAVAPCRAEKLMCSSSRS  212

Scoring table: OLIGO
             Xgapop   60.0 , Xgapext   60.0
             Ygapop   60.0 , Ygapext   60.0
             Fgapop    6.0 , Fgapext    7.0
             Delop     6.0 , Delext     7.0

Searched:    5883141 segs, 28421725653 residues

Word size:   1

Total number of hits satisfying chosen parameters:      11756320

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095515_17514/app_query.fasta_1
-DB=GenEmbl -QFMT=fastap -SUFFIX=oligo_p2n.rge -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=500 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=100 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=200 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss03h
-USER=US09989890 @CGN_1_1.4939 @runat 16032006_095515_17514 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :   GenEmbl:*
             1: gb_ba:*
             2: gb_in:*
             3: gb_env:*
             4: gb_om:*
             5: gb_ov:*
             6: gb_pat:*
             7: gb_ph:*
             8: gb_pr:*
             9: gb_ro:*
            10: gb_sts:*
            11: gb_sy:*
            12: gb_un:*
            13: gb_vi:*
            14: gb_htg:*
            15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES

              %
Result      Query
No.   Score  Match  Length DB  ID         Description
----  -----  -----  ------ --  --         -----------
1     212    100.0  1197   6   CQ717675   CQ717675 Sequence
2     212    100.0  1679   8   BC033143   BC033143 Homo sapi
3     212    100.0  1713   6   AX644347   AX644347 Sequence
```

No.	c	Score	%	Length	ID	Description
4		212	100.0	140207	AL356390	AL356390 Human DNA
5		129	60.8	111738	AL592304	AL592304 Homo sapi
6		115	54.2	1785	AX747274	AX747274 Sequence
7		115	54.2	1785	AK091952	AK091952 Homo sapi
8		81	38.2	427	AX071660	AX071660 Sequence
9		59	27.8	1977	AX535029	AX535029 Sequence
10		59	27.8	1977	AX535090	AX535090 Sequence
11		12	5.7	1680	BC006890	BC006890 Mus muscu
12	c	11	5.2	68100	AC004961	AC004961 Homo sapi
13	c	11	5.2	170236	AC147052	AC147052 Pan trogl
14		11	5.2	172424	AC146125	AC146125 Pan trogl
15		10	4.7	525	AF175677	AF175677 Neisseria
16	c	10	4.7	525	AF175679	AF175679 Neisseria
17	c	10	4.7	525	AF175681	AF175681 Neisseria
18	c	10	4.7	1166	STMNUSG	D17465 S.coelicolo
19		10	4.7	1487	BC048207	BC048207 Homo sapi
20	c	10	4.7	1635	CQ841901	CQ841901 Sequence
21	c	10	4.7	1635	AK124962	AK124962 Homo sapi
22	c	10	4.7	1718	BC079148	BC079148 Rattus no
23		10	4.7	1968	BT007455	BT007455 Homo sapi
24		10	4.7	1968	AY890173	AY890173 Synthetic
25		10	4.7	1975	AY890174	AY890174 Synthetic
26	c	10	4.7	2162	STMSEC	D17464 S.griseus g
27		10	4.7	2413	AY094061	AY094061 Sus scrof
28		10	4.7	2482	BC021981	BC021981 Homo sapi
29		10	4.7	2816	BC070065	BC070065 Homo sapi
30		10	4.7	2816	CQ843453	CQ843453 Sequence
31	c	10	4.7	2981	AK126430	AK126430 Homo sapi
32		10	4.7	2981	STMSAM	D32254 Streptomyce
33		10	4.7	3421	CQ874011	CQ874011 Sequence
34		10	4.7	3421	HSU02310	U02310 Human fork
35		10	4.7	5723	CS031181	CS031181 Sequence
36		10	4.7	5723	CS040133	CS040133 Sequence
37		10	4.7	5723	AX587664	AX587664 Sequence
38		10	4.7	5723	AF032885	AF032885 Homo sapi
39		10	4.7	5769	AX779990	AX779990 Sequence
40		10	4.7	7235	SGNUSG	X72787 S.griseus n
41	c	10	4.7	68170	AC137577	AC137577 Homo sapi
42		10	4.7	69695	BA000002_16	Continuation (17 o
43		10	4.7	110000	BA000030_59	Continuation (60 o
44	c	10	4.7	110000	CR555291_0	CR555291 Danio rer
45		10	4.7	143701	HS29C18	Z97192 Human DNA
46		10	4.7	151261	AL355132	AL355132 Human DNA
47		10	4.7	156763	AC005901	AC005901 Homo sapi
48		10	4.7	174869	CR936217	CR936217 Danio rer
49		10	4.7	175626	CR846094	CR846094 Zebrafish
50	c	10	4.7	199160	AC137538	AC137538 Canis fam
51	c	10	4.7	210174	AC124555	AC124555 Mus muscu
52		10	4.7	216621	AC158593	AC158593 Mus muscu
53		10	4.7	217987	CR792443	CR792443 Zebrafish
54		10	4.7	226168	AC078930	AC078930 Mus muscu
55	c	10	4.7	235419	AC095979	AC095979 Rattus no
56		10	4.7	248550	SCO939120	AL939120 Streptomy
57		10	4.7	249406	AC118963	AC118963 Rattus no
58	c	10	4.7	254687	AC137427	AC137427 Rattus no
59		10	4.7	256017	AC126312	AC126312 Rattus no
60		10	4.7	346316	AC107549	AC107549 Rattus no
61		9	4.2	100	AB032303	AB032303 TT virus
62		9	4.2	100	AB032304	AB032304 TT virus
63		9	4.2	140	HSZ95374	Z95374 H.sapiens g
64		9	4.2	303	AY745220	AY745220 Anopheles
65	c	9	4.2	340	BV106213	BV106213 MARC 3519
66	c	9	4.2	354	G67666	G67666 MARC3174-31
67		9	4.2	354	BV104173	BV104173 MARC 3174
68		9	4.2	379	BV104270	BV104270 MARC 3198
69		9	4.2	387	AF402268	AF402268 Anaplasma
70		9	4.2	442	DM190E7T	AL023365 Drosophil
71	c	9	4.2	453	DM96A7T	AL023471 Drosophil
72	c	9	4.2	501	BV308315	BV308315 S236P6209
73	c	9	4.2	546	BD078483	BD078483 101 human
74	c	9	4.2	546	CQ849043	CQ849043 Sequence
75	c	9	4.2	604	BV345691	BV345691 S230P6315
76		9	4.2	1119	CS074643	CS074643 Sequence

c	#			len	n	ID	Accession	Description
	77	9	4.2	1210	15	AY150443	AY150443	Arabidops
	78	9	4.2	1301	15	AY508007	AY508007	Helianthu
	79	9	4.2	1327	15	AY188374	AY188374	Sporisori
	80	9	4.2	1362	8	BC101317	BC101317	Homo sapi
c	81	9	4.2	1377	6	AR626918	AR626918	Sequence
	82	9	4.2	1436	15	AY074356	AY074356	Arabidops
	83	9	4.2	1448	6	AF153341	AF153341	Homo sapi
c	84	9	4.2	1644	6	AX685724	AX685724	Sequence
	85	9	4.2	1644	8	AY437532	AY437532	Homo sapi
c	86	9	4.2	1716	6	AR627633	AR627633	Human DNA
c	87	9	4.2	1761	15	D89272	D89272	Schizosacch
c	88	9	4.2	1818	6	AX687001	AX687001	Sequence
	89	9	4.2	1912	6	CQ876112	CQ876112	Sequence
c	90	9	4.2	2225	6	AX687002	AX687002	Sequence
	91	9	4.2	2322	8	BC053850	BC053850	Homo sapi
c	92	9	4.2	2386	4	BT021625	BT021625	Bos tauru
	93	9	4.2	2417	5	BC085632	BC085632	Danio rer
c	94	9	4.2	2576	15	AK073176	AK073176	Oryza sat
	95	9	4.2	2893	8	HUMAGCGB	L13434	Human chrom
c	96	9	4.2	3353	15	AK065775	AK065775	Oryza sat
	97	9	4.2	3699	8	BC101314	BC101314	Homo sapi
c	98	9	4.2	3840	6	CQ596190	CQ596190	Sequence
	99	9	4.2	3956	9	MUSRRPB	M26651	Mouse rearr
	100	9	4.2	4109	8	AF225950	AF225950	Homo sapi
	101	9	4.2	5906	6	CQ596189	CQ596189	Sequence
	102	9	4.2	6495	13	ASG241KP	D14995	Apple stem
	103	9	4.2	7874	6	AR619473	AR619473	Sequence
	104	9	4.2	9198	6	AR619571	AR619571	Sequence
	105	9	4.2	9455	15	AF480446	AF480446	Ustilago
	106	9	4.2	9839	9	AB045007	AB045007	Mus muscu
	107	9	4.2	10029	1	AE015576	AE015576	Shewanell
c	108	9	4.2	10432	6	AR447383	AR447383	Sequence
c	109	9	4.2	10432	6	AR447466	AR447466	Sequence
c	110	9	4.2	10432	8	AR531514	AR531514	Sequence
c	111	9	4.2	11435	6	CQ728457	CQ728457	Sequence
c	112	9	4.2	11435	6	CS032994	CS032994	Sequence
c	113	9	4.2	11435	8	CS041946	CS041946	Sequence
c	114	9	4.2	11435	8	HSMKI67A	X65551	H.sapiens m
c	115	9	4.2	11971	14	AC017863	AC017863	Drosophil
c	116	9	4.2	12493	6	AX009576	AX009576	Sequence
c	117	9	4.2	12498	10	BV176969	BV176969	sqm91895
c	118	9	4.2	12498	10	BV176970	BV176970	sqm91897
c	119	9	4.2	12515	6	CQ981425	CQ981425	Sequence
c	120	9	4.2	12515	6	CS032990	CS032990	Sequence
c	121	9	4.2	12515	6	CS032992	CS032992	Sequence
c	122	9	4.2	12515	6	CS041942	CS041942	Sequence
c	123	9	4.2	12515	6	CS041944	CS041944	Sequence
c	124	9	4.2	12515	6	CS052349	CS052349	Sequence
c	125	9	4.2	12515	6	CS063108	CS063108	Sequence
c	126	9	4.2	12515	6	CS080889	CS080889	Sequence
c	127	9	4.2	12515	6	AX336950	AX336950	Sequence
c	128	9	4.2	12516	8	HSMKI67	X65550	H.sapiens m
c	129	9	4.2	12516	10	BV175100	BV175100	sqm76691
c	130	9	4.2	12747	6	AX780122	AX780122	Sequence
c	131	9	4.2	12747	6	AX780123	AX780123	Sequence
c	132	9	4.2	15451	2	TRE245743	AJ245743	Terebratu
	133	9	4.2	18979	9	BX842567	BX842567	Mouse DNA
	134	9	4.2	29111	8	AJ698723	AJ698723	Stigmatel
	135	9	4.2	29965	8	HSAMAK67	X74107	H.sapiens g
	136	9	4.2	37342	15	SPBC337	AL031854	S.pombe c
	137	9	4.2	38381	2	AC000039	AC000039	Homo sapi
	138	9	4.2	38397	8	DMC8D8	AL022018	Drosophil
	139	9	4.2	48077	14	AC135043	AC135043	Rattus no
c	140	9	4.2	49275	6	CS024257	CS024257	Sequence
	141	9	4.2	56452	14	AC166273	AC166273	Bos tauru
	142	9	4.2	62761	8	AL499606	AL499606	Human DNA
	143	9	4.2	70875	8	HSD7875H3	AL049733	Human DNA
c	144	9	4.2	71359	14	AC101629	AC101629	Mus muscu
c	145	9	4.2	73544	6	AX685726	AX685726	Sequence
c	146	9	4.2	74969	14	AC016269	AC016269	Homo sapi
c	147	9	4.2	77858	3	AY688432	AY688432	Uncidentured
c	148	9	4.2	78210	1	AB070949	AB070949	Streptomy
c	149	9	4.2	83370	14	AP008188	AP008188	Lotus cor

c	#			len	n	ID	Accession	Description
c	150	9	4.2	86381	14	AL360077	AL360077	Homo sapi
c	151	9	4.2	87351	14	AL365316	AL365316	Homo sapi
	152	9	4.2	87563	8	AC091601	AC091601	Homo sapi
	153	9	4.2	89203	14	AC119504	AC119504	Rattus no
c	154	9	4.2	91869	8	AL353773	AL353773	Human DNA
c	155	9	4.2	92118	8	AC074384	AC074384	Homo sapi
c	156	9	4.2	92652	15	AC024261	AC024261	Arabidops
c	157	9	4.2	93650	14	AL359635_4		Continuation (5 of
c	158	9	4.2	96600	6	AX695710	AX695710	Sequence
c	159	9	4.2	98593	8	AL356472	AL356472	Human DNA
	160	9	4.2	104171	14	AC019499	AC019499	Drosophil
c	161	9	4.2	110000	1	CP000091_09		Continuation (10 o
	162	9	4.2	110000	1	AP006618_33		Continuation (34 o
	163	9	4.2	110000	1	AP006627_32		Continuation (33 o
c	164	9	4.2	110000	1	AY596297_11		Continuation (12 o
c	165	9	4.2	110000	1	BA000012_03		Continuation (4 of
	166	9	4.2	110000	1	BA000012_33		Continuation (34 o
	167	9	4.2	110000	1	BA000030_04		Continuation (5 of
	168	9	4.2	110000	1	BA000030_05		Continuation (6 of
c	169	9	4.2	110000	1	BA000035_19		Continuation (20 o
c	170	9	4.2	110000	1	CP000076_53		Continuation (54 o
	171	9	4.2	110000	14	CT005247_2		Continuation (3 of
	172	9	4.2	110000	14	CT005263_3		Continuation (4 of
	173	9	4.2	110000	15	AP008214_241		Continuation (242
	174	9	4.2	110000	15	AP008215_030		Continuation (31 o
	175	9	4.2	110000	15	AP008216_049		Continuation (50 o
	176	9	4.2	110000	15	AP008217_204		Continuation (205
	177	9	4.2	110000	15	AP008218_232		Continuation (233
	178	9	4.2	110000	15	AP008207_410		Continuation (411
	179	9	4.2	110000	15	AP008210_334		Continuation (335
c	180	9	4.2	110000	15	AP008211_227		Continuation (228
	181	9	4.2	116821	8	AL606537	AL606537	Human DNA
	182	9	4.2	116951	8	AC114788	AC114788	Homo sapi
	183	9	4.2	119535	8	AL606446	AL606446	Oryza sat
	184	9	4.2	121768	15	AP005560	AP005560	Oryza sat
	185	9	4.2	122888	8	AC022326	AC022326	Human DNA
	186	9	4.2	123057	8	HS333H23	AL390236	Human DNA
	187	9	4.2	125086	8	AL607044	AL607044	Human DNA
	188	9	4.2	125506	14	AC160611	AC160611	Loxodonta
	189	9	4.2	125685	8	AL513263	AL513263	Human DNA
	190	9	4.2	128090	8	AC073310	AC073310	Homo sapi
	191	9	4.2	129689	14	AC108372	AC108372	Pan trogl
	192	9	4.2	129689	15	AC140550	AC140550	Human DNA
	193	9	4.2	133961	8	AL136454	AL136454	Human DNA
	194	9	4.2	136068	14	AC161732	AC161732	Loxodonta
	195	9	4.2	137676	15	AC137693	AC137693	Oryza sat
	196	9	4.2	139163	14	AC147763	AC147763	Dasypus n
	197	9	4.2	140466	14	AC148750	AC148750	Dasypus n
	198	9	4.2	141535	8	AL355529	AL355529	Human DNA
	199	9	4.2	142239	15	AC108503	AC108503	Oryza sat
	200	9	4.2	144433	15	AP003266	AP003266	Oryza sat
	201	9	4.2	146099	14	AC152466	AC152466	Dasypus n
	202	9	4.2	147533	14	AC011192	AC011192	Homo sapi
	203	9	4.2	148213	8	AP001979	AP001979	Homo sapi
	204	9	4.2	148301	9	AL645731	AL645731	Mouse DNA
	205	9	4.2	150818	15	AC098566	AC098566	Oryza sat
	206	9	4.2	152959	15	AP004557	AP004557	Oryza sat
	207	9	4.2	153108	14	OSJN00090	OSJN00090	Oryza sat
	208	9	4.2	153214	15	AC142163	AC142163	Rattus no
	209	9	4.2	153520	14	CR956366	CR956366	Sus scrof
	210	9	4.2	153670	14	AC161732	AC161732	Oryza sat
	211	9	4.2	154016	15	AC121361	AC121361	Oryza sat
	212	9	4.2	154732	15	AP005727	AP005727	Oryza sat
	213	9	4.2	158438	8	AL355295	AL355295	Homo sapi
	214	9	4.2	160077	14	AC155716	AC155716	Mus muscu
	215	9	4.2	160602	9	AC011700	AC011700	Homo sapi
	216	9	4.2	160931	8	AC068757	AC068757	Homo sapi
	217	9	4.2	160945	14	AC149568	AC149568	Papio anu
	218	9	4.2	160945	9	AC127299	AC127299	Mus muscu
	219	9	4.2	162568	8	AC093498	AC093498	Drosophil
	220	9	4.2	162568	8	AP003420	AP003420	Homo sapi
c	221	9	4.2	163315	15	AC136149	AC136149	Oryza sat
c	222	9	4.2	164252	2	AC104288	AC104288	Drosophil

Note: AC163955 Oryctolag appears at entry 221 area (AC163955).

flag	idx			pos	cnt	acc	description
c	223	4.2	9	164513	14	AC148212	AC148212 Dasypus n
c	224	4.2	9	165984	14	AC154399	AC154399 Mus muscu
	225	4.2	9	166000	8	AC090509	AC090509 Homo sapi
c	226	4.2	9	166143	14	AL356960	AL356960 Homo sapi
c	227	4.2	9	166294	14	AC122113	AC122113 Atelerix
c	228	4.2	9	166735	14	AC157751	AC157751 Callithri
c	229	4.2	9	167735	14	AC009794	AC009794 Homo sapi
c	230	4.2	9	167970	14	AC156766	AC156766 Rhinoloph
	231	4.2	9	168063	15	CNS07YQ8	AL713908 Oryza sat
c	232	4.2	9	168491	14	AC160919	AC160919 Papio ham
	233	4.2	9	168672	14	AC027102	AC027102 Homo sapi
	234	4.2	9	169391	14	AC111190	AC111190 Homo sapi
c	235	4.2	9	169771	14	AC141890	AC141890 Mus muscu
c	236	4.2	9	170365	9	AL390920	AL390920 Human DNA
	237	4.2	9	171260	14	AC084351	AC084351 Homo sapi
c	238	4.2	9	171718	14	AC141444	AC141444 Homo sapi
	239	4.2	9	172692	14	AP001013	AP001013 Homo sapi
c	240	4.2	9	173716	9	AC107237	AC107237 Mus muscu
c	241	4.2	9	173926	14	MM454K24	AL160232 Mus muscu
	242	4.2	9	174221	8	AC011298	AC011298 Homo sapi
	243	4.2	9	175407	15	AP004620	AP004620 Oryza sat
c	244	4.2	9	175656	14	AC159641	AC159641 Mus muscu
	245	4.2	9	176584	14	AC027812	AC027812 Homo sapi
	246	4.2	9	177626	8	AC090005	AC090005 Homo sapi
c	247	4.2	9	178168	14	AC012350	AC012350 Homo sapi
c	248	4.2	9	178222	9	AC154430	AC154430 Mus muscu
	249	4.2	9	178478	14	AP002834	AP002834 Mus muscu
c	250	4.2	9	179186	14	AL357113	AL357113 Homo sapi
	251	4.2	9	179934	8	AP000810	AP000810 Homo sapi
c	252	4.2	9	180213	14	AC150812	AC150812 Callithri
c	253	4.2	9	181360	2	AC104141	AC104141 Drosophil
c	254	4.2	9	181394	8	AC090951	AC090951 Homo sapi
c	255	4.2	9	181801	14	AC141809	AC141809 Apis mell
	256	4.2	9	182152	14	AC067929	AC067929 Homo sapi
	257	4.2	9	182195	14	AC140431	AC140431 Mus muscu
c	258	4.2	9	182713	14	CR354392	CR354392 Danio rer
	259	4.2	9	183046	6	CS086327	CS086327 Sequence
c	260	4.2	9	183099	14	AC034144	AC034144 Homo sapi
c	261	4.2	9	183330	14	AC121535	AC121535 Mus muscu
c	262	4.2	9	183607	8	AC066597	AC066597 Homo sapi
c	263	4.2	9	183842	14	AC130450	AC130450 Homo sapi
c	264	4.2	9	183991	14	CR926126	CR926126 Danio rer
c	265	4.2	9	184073	5	AL954703	AL954703 Zebrafish
	266	4.2	9	184223	8	HSA261N11	AL121827 Human DNA
c	267	4.2	9	185139	8	BX511254	BX511254 Zebrafish
c	268	4.2	9	185174	9	AC136004	AC136004 Mus muscu
	269	4.2	9	188636	8	HUAC002287	AC002287 Homo sapi
c	270	4.2	9	188791	9	AC117185	AC117185 Mus muscu
c	271	4.2	9	188977	9	AC101658	AC101658 Mus muscu
	272	4.2	9	189286	8	AC008125	AC008125 Homo sapi
c	273	4.2	9	189703	9	AL672119	AL672119 Mouse DNA
c	274	4.2	9	190317	14	AC084332	AC084332 Homo sapi
	275	4.2	9	191058	4	AC150542	AC150542 Bos tauru
c	276	4.2	9	192561	14	AC099532	AC099532 Bos tauru
c	277	4.2	9	193813	9	AL627228	AL627228 Mouse DNA
c	278	4.2	9	193847	9	AC139757	AC139757 Mus muscu
c	279	4.2	9	193863	14	AC149550	AC149550 Callithri
	280	4.2	9	194930	14	AC155497	AC155497 Zea mays
c	281	4.2	9	196451	8	AC112784	AC112784 Homo sapi
c	282	4.2	9	196724	9	AL590963	AL590963 Mouse DNA
c	283	4.2	9	196912	14	AL513346	AL513346 Mouse DNA
	284	4.2	9	197202	9	AC153077	AC153077 Callithri
c	285	4.2	9	197796	14	AC152610	AC152610 Bos tauru
c	286	4.2	9	198241	14	AC150857	AC150857 Bos tauru
c	287	4.2	9	198957	14	AC152226	AC152226 Bos tauru
c	288	4.2	9	199446	14	AC157432	AC157432 Callithri
	289	4.2	9	201478	9	AC132608	AC132608 Mus muscu
c	290	4.2	9	203598	14	AC090250	AC090250 Homo sapi
c	291	4.2	9	204550	14	AC159392	AC159392 Bos tauru
c	292	4.2	9	205690	14	AC159462	AC159462 Mus muscu
c	293	4.2	9	206098	14	AC156915	AC156915 Bos tauru
	294	4.2	9	206957	9	AC091682	AC091682 Mus muscu
c	295	4.2	9	207070	14	CT009632	CT009632 Mus muscu

flag	idx			pos	cnt	acc	description
c	296	4.2	9	207591	14	CR956405	CR956405 Sus scrof
	297	4.2	9	207598	9	AC122901	AC122901 Mus muscu
c	298	4.2	9	207795	14	AC141502	AC141502 Rattus no
c	299	4.2	9	210723	8	AP001858	AP001858 Homo sapi
c	300	4.2	9	210833	14	AC018566	AC018566 Homo sapi
c	301	4.2	9	212060	14	AC148416	AC148416 Callithri
	302	4.2	9	213140	14	AC040911	AC040911 Homo sapi
	303	4.2	9	213519	9	AL714024	AL714024 Mouse DNA
	304	4.2	9	213772	5	AC158116	AC158116 Gallus ga
	305	4.2	9	215126	14	AC083916	AC083916 Mus muscu
	306	4.2	9	215903	14	AC160644	AC160644 Gallus ga
	307	4.2	9	217496	8	AC068319	AC068319 Homo sapi
c	308	4.2	9	217586	9	AC153630	AC153630 Mus muscu
	309	4.2	9	218242	14	AC158678	AC158678 Mus muscu
	310	4.2	9	220807	14	AC105680	AC105680 Rattus no
c	311	4.2	9	220807	14	AC121311	AC121311 Mus muscu
	312	4.2	9	224686	14	AC139643	AC139643 Rattus no
	313	4.2	9	226725	14	AC110974	AC110974 Rattus no
c	314	4.2	9	227248	9	AC125762	AC125762 Rattus no
	315	4.2	9	229563	9	BX001010	BX001010 Mouse DNA
	316	4.2	9	230362	9	AC024957	AC024957 Mus muscu
c	317	4.2	9	230480	14	AC097942	AC097942 Rattus no
	318	4.2	9	230871	14	AC127111	AC127111 Rattus no
	319	4.2	9	231259	14	AC163361	AC163361 Mus muscu
c	320	4.2	9	233361	14	AC160999	AC160999 Mus muscu
	321	4.2	9	234240	9	AC105543	AC105543 Rattus no
c	322	4.2	9	234331	14	AC148018	AC148018 Mus muscu
c	323	4.2	9	236579	14	AC116267	AC116267 Rattus no
	324	4.2	9	237001	9	AC121707	AC121707 Rattus no
c	325	4.2	9	237001	14	AC120124	AC120124 Mus muscu
c	326	4.2	9	239102	14	AC137051	AC137051 Rattus no
	327	4.2	9	239308	14	AC107141	AC107141 Rattus no
	328	4.2	9	240216	14	AC117121	AC117121 Rattus no
	329	4.2	9	241154	14	AC160800	AC160800 Bos tauru
	330	4.2	9	243131	14	AC112421	AC112421 Rattus no
	331	4.2	9	243851	14	AC123448	AC123448 Rattus no
	332	4.2	9	246381	14	AC132545	AC132545 Rattus no
c	333	4.2	9	246522	14	AC096245	AC096245 Rattus no
	334	4.2	9	246793	14	AC160254	AC160254 Bos tauru
c	335	4.2	9	250742	14	AC112102	AC112102 Rattus no
	336	4.2	9	251322	14	AC157234	AC157234 Bos tauru
	337	4.2	9	252595	14	AC152607	AC152607 Bos tauru
c	338	4.2	9	252843	14	AC096261	AC096261 Rattus no
	339	4.2	9	253765	14	AC133612	AC133612 Rattus no
c	340	4.2	9	257003	9	AC122830	AC122830 Mus muscu
c	341	4.2	9	259473	14	AC129868	AC129868 Rattus no
c	342	4.2	9	264050	1	AL627279	AL627279 Salmonell
	343	4.2	9	267357	14	AC115424	AC115424 Rattus no
	344	4.2	9	267481	14	AC114471	AC114471 Rattus no
	345	4.2	9	267809	14	AC137461	AC137461 Rattus no
	346	4.2	9	271686	14	AC126856	AC126856 Rattus no
	347	4.2	9	274723	14	AC161064	AC161064 Bos tauru
	348	4.2	9	277363	14	AC134745	AC134745 Rattus no
	349	4.2	9	277986	14	AC152323	AC152323 Bos tauru
c	350	4.2	9	291278	14	AC097878	AC097878 Rattus no
	351	4.2	9	296820	9	AF312033	AF312033 Mus muscu
c	352	4.2	9	300209	15	AB017064	AB017064 Oryza sat
	353	4.2	9	300592	1	AE016846	AE016846 Salmonell
	354	4.2	9	301817	14	AC095571	AC095571 Rattus no
	355	4.2	9	303314	2	AB003538	AB003538 Drosophil
	356	4.2	9	307462	14	AC112828	AC112828 Rattus no
c	357	3.8	8	308311	1	AE003420	AE003420 Drosophil
	358	3.8	8	340750	1	BX294135	BX294135 Pirellula
	359	3.8	8	55	6	AR014642	AR014642 Sequence
c	360	3.8	8	55	6	AR014643	AR014643 Sequence
	361	3.8	8	55	6	AR072168	AR072168 Sequence
c	362	3.8	8	55	6	AR072169	AR072169 Sequence
	363	3.8	8	55	6	CQ775548	CQ775548 Sequence
c	364	3.8	8	55	6	CQ775549	CQ775549 Sequence
	365	3.8	8	55	6	AR235444	AR235444 Sequence
c	366	3.8	8	55	6	AR235445	AR235445 Sequence
	367	3.8	8	55	6	BD010434	BD010434 Chimeric
c	368	3.8	8	55	6	BD010435	BD010435 Chimeric

Mark	No.	Num	N	E	S	Accession	Description
c	369	75	6	3.8	8	CQ771466	CQ771466 Sequence
c	370	121	10	3.8	8	BV181380	BV181380 sqnm11752
	371	131	6	3.8	8	CQ976587	CQ976587 Sequence
	372	143	6	3.8	8	CQ976586	CQ976586 Sequence
	373	143	6	3.8	8	CQ976588	CQ976588 Sequence
	374	180	6	3.8	8	BD031923	BD031923 Sequence
	375	180	6	3.8	8	AX896390	AX896390 Mus muscu
	376	186	6	3.8	8	CQ079164	CQ079164 Sequence
	377	186	6	3.8	8	CQ110623	CQ110623 Sequence
	378	186	6	3.8	8	CQ149354	CQ149354 Sequence
	379	186	6	3.8	8	CQ184263	CQ184263 Sequence
	380	186	6	3.8	8	CQ232630	CQ232630 Sequence
	381	186	6	3.8	8	CQ270673	CQ270673 Sequence
	382	186	6	3.8	8	CQ307944	CQ307944 Sequence
	383	186	6	3.8	8	CQ344772	CQ344772 Sequence
	384	192	6	3.8	8	CS047765	CS047765 Sequence
	385	201	6	3.8	8	CQ926773	CQ926773 Sequence
	386	201	10	3.8	8	BV173524	BV173524 sqnm62273
	387	201	10	3.8	8	BV198458	BV198458 sqnm19767
	388	201	10	3.8	8	BV204003	BV204003 sqnm21373
	389	257	8	3.8	8	HSDNASEAD	Z22344 H.sapiens D
	390	274	8	3.8	8	BD059815	BD059815 Secreted
	391	303	8	3.8	8	HUMLAMPAO2	L08577 Homo sapien
	392	314	8	3.8	8	HS33C6R	Z65329 H.sapiens C
	393	340	8	3.8	8	CS052239	CS052239 Sequence
	394	340	6	3.8	8	CS112447	CS112447 Sequence
	395	340	8	3.8	8	CS000982	CS000982 Sequence
	396	340	8	3.8	8	HSPCDK1	U12818 Homo sapien
	397	346	8	3.8	8	HSU92987	U92987 Homo sapien
	398	349	8	3.8	8	AX337141	AX337141 Sequence
	399	360	4	3.8	8	AB162851	AB162851 Canis fam
	400	361	5	3.8	8	PRRN28C1	PRRN28C1 Sequence
	401	370	5	3.8	8	SCRN28C1	SCRN28C1 Sequence
	402	375	5	3.8	8	ECRN28C1	Z18676 Erpetoichth
	403	382	6	3.8	8	AX330678	AX330678 Sequence
	404	382	6	3.8	8	AX330902	AX330902 Sequence
	405	382	6	3.8	8	AX334084	AX334084 Sequence
	406	382	6	3.8	8	AX410327	AX410327 Sequence
	407	382	10	3.8	8	G71067	G71067 A62219334FB
c	408	383	15	3.8	8	AJ010431	AJ010431 Hylocomiu
	409	384	5	3.8	8	CARN28C1	CARN28C1 Sequence
	410	395	6	3.8	8	CS010357	CS010357 Sequence
	411	406	6	3.8	8	CS118561	CS118561 Sequence
	412	411	6	3.8	8	CQ407602	CQ407602 Sequence
	413	418	10	3.8	8	G71936	G71936 A39183134FM
	414	428	6	3.8	8	CQ112591	CQ112591 Sequence
	415	428	6	3.8	8	CQ151462	CQ151462 Sequence
	416	428	6	3.8	8	CQ185238	CQ185238 Sequence
	417	428	6	3.8	8	CQ234845	CQ234845 Sequence
	418	428	6	3.8	8	CQ272396	CQ272396 Sequence
	419	428	6	3.8	8	CQ310088	CQ310088 Sequence
	420	428	6	3.8	8	CQ346675	CQ346675 Sequence
	421	432	5	3.8	8	AF204994	AF204994 Coccyzus
	422	436	6	3.8	8	CQ507214	CQ507214 Sequence
	423	436	10	3.8	8	G71937	G71937 A39185834FM
c	424	441	6	3.8	8	AR176419	AR176419 Sequence
	425	441	6	3.8	8	BD226032	BD226032 Compound
	426	441	6	3.8	8	BD275703	BD275703 COMPOUNDS
	427	441	6	3.8	8	AR220488	AR220488 Sequence
	428	441	8	3.8	8	AR255482	AR255482 Sequence
	429	441	6	3.8	8	AR281052	AR281052 Sequence
	430	441	6	3.8	8	AR437843	AR437843 Sequence
	431	441	6	3.8	8	AR476379	AR476379 Sequence
	432	441	6	3.8	8	AR486570	AR486570 Sequence
	433	441	6	3.8	8	AR541073	AR541073 Sequence
	434	441	6	3.8	8	AR606460	AR606460 Sequence
	435	441	6	3.8	8	AX365704	AX365704 Sequence
c	436	442	6	3.8	8	CQ069945	CQ069945 Sequence
c	437	442	6	3.8	8	CQ097451	CQ097451 Sequence
	438	442	6	3.8	8	CQ136283	CQ136283 Sequence
	439	442	6	3.8	8	CQ174355	CQ174355 Sequence
c	440	442	6	3.8	8	CQ219583	CQ219583 Sequence
c	441	442	6	3.8	8	CQ258067	CQ258067 Sequence
c	442	442	6	3.8	8	CQ295390	CQ295390 Sequence
c	443	442	6	3.8	8	CQ331994	CQ331994 Sequence
	444	448	10	3.8	8	G58617	G58617 SHGC-104639
	445	449	4	3.8	8	AY818354	AY818354 Sus scrof
c	446	456	10	3.8	8	G52741	G52741 SHGC-79560
	447	464	9	3.8	8	CQ483103	CQ483103 Sequence
	448	464	9	3.8	8	BC059725	BC059725 Mus muscu
c	449	465	5	3.8	8	AY147973	AY147973 Rana temp
	450	473	6	3.8	8	AX172561	AX172561 Sequence
	451	479	6	3.8	8	AR501292	AR501292 Sequence
	452	479	6	3.8	8	AR516574	AR516574 Sequence
	453	481	6	3.8	8	AF305505	AF305505 Anaplasma
	454	481	10	3.8	8	G71965	G71965 A62219334FM
c	455	488	6	3.8	8	AX753039	AX753039 Sequence
c	456	488	6	3.8	8	AX753040	AX753040 Sequence
c	457	492	6	3.8	8	AX753035	AX753035 Sequence
	458	493	6	3.8	8	CQ484322	CQ484322 Sequence
	459	506	6	3.8	8	CQ752871	CQ752871 Sequence
	460	509	6	3.8	8	AX261627	AX261627 Sequence
	461	522	5	3.8	8	BC054686	BC054686 Danio rer
	462	523	4	3.8	8	DQ127247	DQ127247 Canis fam
	463	525	1	3.8	8	AF175676	AF175676 Neisseria
	464	525	1	3.8	8	AF175678	AF175678 Neisseria
	465	525	1	3.8	8	AF175680	AF175680 Neisseria
	466	525	1	3.8	8	AF175682	AF175682 Neisseria
	467	525	1	3.8	8	AF175683	AF175683 Neisseria
	468	525	1	3.8	8	AY157539	AY157539 Neisseria
	469	525	6	3.8	8	CQ786532	CQ786532 Sequence
	470	525	6	3.8	8	CQ814512	CQ814512 Sequence
	471	525	6	3.8	8	CQ920533	CQ920533 Sequence
	472	526	6	3.8	8	AX778165	AX778165 Sequence
	473	528	6	3.8	8	AR167420	AR167420 Sequence
	474	535	2	3.8	8	AB004766	AB004766 Bombyx mo
	475	543	6	3.8	8	AX778486	AX778486 Sequence
	476	546	15	3.8	8	AK105218	AK105218 Oryza sat
	477	561	1	3.8	8	AF305504	AF305504 Anaplasma
	478	567	6	3.8	8	BD136041	BD136041 Interfero
	479	569	6	3.8	8	AR659436	AR659436 Sequence
	480	571	10	3.8	8	BV396525	BV396525 S243P681R
c	481	572	6	3.8	8	CQ921784	CQ921784 Sequence
	482	577	6	3.8	8	CQ099510	CQ099510 Sequence
	483	577	6	3.8	8	CQ138494	CQ138494 Sequence
	484	577	6	3.8	8	CQ175359	CQ175359 Sequence
	485	577	6	3.8	8	CQ221915	CQ221915 Sequence
	486	577	6	3.8	8	CQ259855	CQ259855 Sequence
	487	577	6	3.8	8	CQ297654	CQ297654 Sequence
	488	577	6	3.8	8	CQ333981	CQ333981 Sequence
	489	578	6	3.8	8	AR148043	AR148043 Sequence
	490	578	6	3.8	8	BD084418	BD084418 Compositi
	491	578	6	3.8	8	AR219767	AR219767 Sequence
	492	578	6	3.8	8	AR350865	AR350865 Sequence
	493	578	6	3.8	8	AR433241	AR433241 Sequence
	494	578	6	3.8	8	AR614985	AR614985 Sequence
	495	578	6	3.8	8	AR642027	AR642027 Sequence
	496	578	6	3.8	8	AX316885	AX316885 Sequence
c	497	578	10	3.8	8	BV285800	BV285800 S232P6524
c	498	579	10	3.8	8	BV291660	BV291660 S232P6568
c	499	580	5	3.8	8	AY708173	AY708173 Philautus
	500	582	9	3.8	8	MMU243031	AJ243031 Mus muscu

ALIGNMENTS

RESULT 1
CQ717675

```
LOCUS       CQ717675                1197 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 3609 from Patent WO02068579.
ACCESSION   CQ717675
VERSION     CQ717675.1  GI:42278532
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

REFERENCE   1
AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE   Kits, such as nucleic acid arrays, comprising a majority of
             humanexons or transcripts, for detecting expression and other uses
             thereof
JOURNAL   Patent: WO 02068579-A 3609 06-SEP-2002;
             PE Corporation (NY) (US)

FEATURES   Location/Qualifiers
source   1..1197
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:   1.58e-192   Length: 1197
Score: 212.00   Matches: 212
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x CQ717675 (1-1197)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   104 AGCCCCCACCAAGCCGCCGCCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 163

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   164 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCCCCTTGAGTCCCCACCTGCT 223

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGTGCCCGGCGTCCTCCGCCGCT 283

Qy   61 AlaGlyIleAlaSerArgAlaValAlaGluProValCysGlyAspAlaAlaProAlaCysLeu 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   284 GCGGGAGTGCCTCCAGCGCGTGGAGCCGTGTGTGGGGGATGCAGCCCCTGCCTGTCTA 343

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   344 CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC 403

Qy   101 ProProAlaLeuIleValHisProProAlaGlyMetAlaSerGlySerSerGlnPro 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404 CCCCCAGCCCTGATCGTGCACCCCCCAGCGCGGGGATGGCCAGCGGCTCAAGTCAACCA 463

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysSerAlaSerLeuCysIleProThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   464 TGGGCAGCAGCTTCAGCACCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTATCCCTACC 523

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   524 CGAGGGGCCACCTCCCCAGCCCTGATGCGGGACTCCTGCTGCAAGGAGCCACTGGCCGATC 583

Qy   161 ProHisProCysAspThrAlaCysProAlaProLeuProValLeuValAlaAlaProArg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   584 CCCCACCCATGCAACAGCCTGCCCAGCACCTTTGCCAGTAGTCCTCGTGGCTCCGAGG 643

Qy   181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgTrpProAlaValAlaProCys 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   644 AGTACTATTCTTTCCATGAGTCGACCTGGACCTGCCGGGAGATGGGCAGTGGCTTCCATGT 703

Qy   201 ArgAlaGluGluLysLeuMetCysSerSerSerArgSer 212
          ||||||||||||||||||||||||||||||||||||
Db   704 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAGAAGC 739

RESULT 2
BC033143
LOCUS   BC033143   1679 bp   mRNA   linear   PRI 03-AUG-2005

DEFINITION Homo sapiens chromosome 1 open reading frame 172, mRNA (cDNA clone
            MGC:45763 IMAGE:3960220), complete cds.
ACCESSION   BC033143
VERSION   BC033143.1 GI:21619958
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 1679)
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
CONSRTM   Mammalian Gene Collection Program Team
TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED   12477932
REFERENCE   2 (bases 1 to 1679)
AUTHORS   NIH MGC Project
CONSRTM   NIH MGC Project
TITLE   Direct Submission
JOURNAL   Submitted (25-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Bethesda, MD 20892-2590, USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Martin Hirst, Thomas Zeng, Ryan Morin, Michelle Moksa, Johnson
            Pang, Diana Mah, Jing Wang, Kieth Fichter, Eric Chuah, Allen
            Delaney, Rob Kirkpatrick, Agnes Baross, Sarah Barber, Mabel
            Brown-John, Steve S. Chand, William Chow, Ryan Babakaiff, Dave
            Wong, Corey Matsuo, Jaclyn Beland, Susan Gibson, Luis delRio, Ruth
            Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
            Kim MacDonald, Mike R. Mayo, Josh Moran, Diana Palmquist, JR
            Santos, Duane Smailus, Jeff Stott, Miranda Tsai, George Yang,
            Jacquie Schein, Asim Siddiqui, Steven Jones, Rob Holt, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 43 Row: c Column: 17.
FEATURES   Location/Qualifiers
source   1..1679
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:45763 IMAGE:3960220"
                     /tissue_type="Placenta, choriocarcinoma"
                     /clone_lib="NIH_MGC_21"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
gene   1..1679

DEFINITION Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.

```
CDS                     /gene="C1orf172"
                        /note="synonyms: FLJ34633, RP11-344H11.3"
                        /db_xref="GeneID:126695"
                        387_.1121
                        /gene="C1orf172"
                        /codon_start=1
                        /product="C1orf172 protein"
                        /protein_id="AAH33143.1"
                        /db_xref="GI:21619959"
                        /db_xref="GeneID:126695"
                        /translation="MGSSFSYPDVKLKGIPVYPYPRATSPAPDADSCCKGPLADPPPM
                        RHSLPSTFASSPRGSEBEYYSFHESDLDLPEMGSGSMSSREIDVLIFKKLTELFSVHQI
                        DELAKCTSDTVFLEKTSKISDLLISSITQDYHLDEQDAEGRLVRGIIRISTRKSRARPQ
                        TSEGRSTRAAAPTAAAPDSGHETMVGSGLSQDELTVQISQETTADAIARKLRPYGAPG
                        YPASHDSSFQGTDTDSSGAPILQVYC"

ORIGIN

Alignment Scores:
Pred. No.:              2.11e-192              Length:              1679
Score:                  212.00                 Matches:             212
Percent Similarity:     100.0%                 Conservative:        0
Best Local Similarity:  100.0%                 Mismatches:          0
Query Match:            100.0%                 Indels:              0
DB:                     8                      Gaps:                0
```

```
LOCUS                   AX644347                1713 bp    DNA          linear   PAT 27-FEB-2003
DEFINITION              Sequence 105 from Patent WO02077232.
ACCESSION               AX644347
VERSION                 AX644347.1     GI:28610406
KEYWORDS                .
SOURCE                  Homo sapiens (human)
  ORGANISM              Homo sapiens
                        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                        Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                        Hominidae; Homo.
REFERENCE               1
  AUTHORS               Salceda,S., Macina,R.A., Recipon,H., Pluta,J., Sun,Y. and Liu,C.
  TITLE                 Compositions and methods relating to breast specific genes and
                        proteins
  JOURNAL               Patent: WO 02077232-A 105 03-OCT-2002;
                        Diadexus, Inc. (US)
FEATURES                Location/Qualifiers
     source             1..1713
                        /organism="Homo sapiens"
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              2.15e-192              Length:              1713
Score:                  212.00                 Matches:             212
Percent Similarity:     100.0%                 Conservative:        0
Best Local Similarity:  100.0%                 Mismatches:          0
Query Match:            100.0%                 Indels:              0
DB:                                            Gaps:                6
```

US-09-989-890-238 (1-212) x BC033143 (1-1679)

```
Qy     1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
Db    28 AGCCCCCACCAAGCCGCGCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC  87

Qy    21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla  40
Db    88 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCCCTTGAGTCCCCACCTGCT  147

Qy    41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
Db   148 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCGCCGCCGCCT  207

Qy    61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaAlaProAlaCysLeu  80
Db   208 GCCGGGATTGCCTCCAGCGCTGTGTGCCGGGGATGCAGCCCCTGCTCTGTCTA  267

Qy    81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
Db   268 CTGAGGACTCCACTGAGGGGACTGCTGAAGCCCGGCTCAAGTCAACCA  327

Qy   101 ProProAlaLeuIleValHisProAlaGlyGlyMetAlaSerGlySerGlnPro  120
Db   328 CCCCCAGCCCTGATCGTGCACCCCGCCGGGGGGATGGCCAGCGGCTCAAGTGAGTGC  387

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
Db   388 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC  447

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
Db   448 CGAGGGGCCACCTCCCCAGCCCCTGATGCCGACTCCTGCTCAAGGAGCCACTGGCCGATC  507

Qy   161 ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProCys  180
Db   508 CCCCACCCATGCCGACACAGCCTGCCCCAGCACCTTTGCCAGTGTCCTGTGGCCGCCGATC  567

Qy   181 SerThrIleLeuSerMetSerArgThrThrCysArgArgTrpAlaValAlaProCys  200
Db   568 AGTACTATTCTTTCCATGAGTCCGACCTGCCGAGATGGGCAGTGGGCTCCATGT  627

Qy   201 ArgAlaGluLysMetCysSerSerSerArgSer  212
Db   628 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC  663

RESULT 3
AX644347
```

US-09-989-890-238 (1-212) x AX644347 (1-1713)

```
Qy     1 SerProHisGlnAlaAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
Db   104 AGCCCCCACCAAGCCGCGCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC  163

Qy    21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla  40
Db   164 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCCCTTGAGTCCCCACCTGCT  223

Qy    41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla  60
Db   224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCGCCGCCGCT  283

Qy    61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaAlaProAlaCysLeu  80
Db   284 GCCGGGATTGCCTCCAGCGCTGTGTGCCGGGGATGCAGCCCCTGCTCTGTCTA  343

Qy    81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
Db   344 CTGAGGACTCCACTGAGGGGACTGCTGAAGCCCGGCTCAAGTGAGCACAATGGAGTGC  403

Qy   101 ProProAlaLeuIleValHisProAlaGlyGlyMetAlaSerGlySerGlnPro  120
Db   404 CCCCCAGCCCTGATCGTGCACCCCGCCGGGGGGATGGCCAGCGGCTCAAGTCAACCA  463

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
Db   464 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC  523

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle  160
Db   524 CGAGGGGCCACCTCCCCAGCCCCTGATGCCGACTCCTGCTCAAGGAGCCACTGGCCGATC  583

Qy   161 ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg  180
Db   584 CCCCACCCATGCCGACACAGCCTGCCCCAGCACCTTTGCCAGTGTCCTGTGGCCGCCGAGG  643

Qy   181 SerThrIleLeuSerMetSerArgThrThrCysArgArgTrpAlaValAlaProCys  200
Db   644 AGTACTATTCTTTCCATGAGTCCGACCTGCCGAGATGGGCAGTGGGCTCCATGT  703
```

```
Qy        201  ArgAlaGluLysMetCysSerSerArgSer 212
               |||||||||||||||||||||||||||||||
Db        704  CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 739

RESULT 4
AL356390
LOCUS       AL356390           140207 bp    DNA        linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-344H11 on chromosome 1 Contains
            the NROB2 gene for nuclear receptor subfamily 0 (group B, member
            2), the 5' end of a novel gene (FLJ12455), the NUDC gene fornuclear
            distribution gene C homolog (A. nidulans), a novel gene (FLJ34633),
            a ribosomal protein L12 (RPL12) pseudogene, a novel protein L32
            (RPL32) pseudogene, a novel pseudogene (DC2), two novel genes and
            two CpG islands, complete sequence.
ACCESSION   AL356390
VERSION     AL356390.24  GI:21435506
KEYWORDS    HTG; DC2; FLJ12455; FLJ34633; NROB2; NUDC; RPL12; RPL32.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 140207)
  AUTHORS   Glithero,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 17, 2002 this sequence version replaced gi:20218554.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-344H11 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            ----------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            ----------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1..140207
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-344H11"
                     /clone_lib="RPCI-11.2"
     misc_feature    496..726
                     /note="Sequence from overlapping clone RP4-633N17
                     (AL137860). Assembly confirmed by restriction digest."
     misc_feature    898..1641
                     /note="Sequence from overlapping clone RP4-633N17
                     (AL137860). Assembly confirmed by restriction digest."
     misc_feature    2000
                     /note="Clone_right_end: RP4-633N17"
     gene            join(21468..22032,27482..28803,29251..29417)
                     /gene="RP11-344H11.8"
     mRNA            join(21468..22032,27482..28803,29251..29417)
                     /gene="RP11-344H11.8"
                     /locus_tag="RP11-344H11.8-002"
                     /product="novel protein"
                     /note="match: cDNAs: Em:AK122816.1"
     gene            join(21603..22032,27482..29419)
                     /gene="RP11-344H11.8"
                     /locus_tag="RP11-344H11.8-001"
     mRNA            join(21603..22032,27482..29419)
                     /gene="RP11-344H11.8"
                     /locus_tag="RP11-344H11.8-001"
                     /product="novel protein"
                     /note="match: cDNAs: Em:AK028944.1 Em:BC012790.1
                     Em:BC012790.2 Em:BC014160.1 Em:BC014160.2 Em:BC038001.1"
     CDS             join(21772..22032,27482..28495)
                     /gene="RP11-344H11.8"
                     /locus_tag="RP11-344H11.8-001"
                     /standard_name="OTTHUMP00000004474"
                     /note="match: proteins: Tr:Q8CI52 Tr:Q8CI10 Tr:Q96A09"
                     /codon_start=1
                     /product="novel protein"
                     /protein_id="CAI13555.1"
                     /db_xref="GI:55959148"
                     /db_xref="Genew:28273"
                     /db_xref="UniProt/TrEMBL:Q96A09"
                     /translation="MPSSSGABRRDRAAAQVGTAAATAVATAAPAGGGPDPEALSAFP
                     GRHLSGLSWPQVKRLDALLSEPIPIHGRGNFPTLSVQPRQIVQVVRSTLEEQGLHVHS
                     VRLHGSAASHVLHPESGLGYKDLDLVFRVDLRSEASFQLTKAVVLACLLDFLPAGVSR
                     AKTPLTLKEAYVQKLVKVCTDSDRWSLISLSNKSGKNVELKFVDSVRRQFEFSIDSF
                     QILDSLLLFGQCSSTPMSEAFHPTVTGBSLYGDFTEALEHLHRHVIATRSPEIRGG
                     GLLKYCHLLVRGFRPRPSTDVRALQRYMCSRFFIDFPDLVEQRTLERYLEAHFGGAD
                     AARRYACLVTLHRVVNESTVCLMNHERRQTLDLIAALALQALAEQGPAATAALAWRPP
                     GTDGVVPATVNYVVTPVQPLLAHAYPTWLPCN"
     polyA_signal    29398..29403
                     /gene="RP11-344H11.8"
     polyA_site      29419
                     /gene="RP11-344H11.8"
     gene            complement(join(33541..34592,39827..40731))
                     /locus_tag="RP11-344H11.8-001"
     mRNA            complement(join(33541..34592,39827..40731))
                     /locus_tag="RP11-344H11.7-001"
                     /product="novel transcript"
                     /note="match: ESTs: Em:AI435085.1 Em:BG752663.1
                     Em:BM021923.1 Em:BM682682.1 Em:BQ646324.1 Em:BQ892187.1
                     Em:CA438061.1"
     gene            48099..48542
                     /locus_tag="RP11-344H11.6-001"
                     /pseudo
     CDS             48099..48542
                     /locus_tag="RP11-344H11.6-001"
                     /note="match: proteins: Tr:AAH54857 Tr:BAB28595
                     Tr:EAA12661 Tr:Q7ZWJ3 Tr:Q8TBU1 Tr:Q9CPZ2 Tr:Q9NRP0
                     Tr:Q9P075 Tr:Q9PIR4"
                     /pseudo
                     /codon_start=1
                     /product="novel pseudogene"
                     /db_xref="PSEUDO:CAI13556.1"
     gene            complement(50415..50811)
                     /locus_tag="RP11-344H11.5-001"
                     /pseudo
     CDS             complement(50415..50811)
                     /locus_tag="RP11-344H11.5-001"
                     /note="match: proteins: Sw:O94008 Sw:P02433 Sw:P17932
                     Sw:P38061 Tr:AAH46339 Tr:AAP80706 Tr:BAC21646 Tr:BAC25812
                     Tr:CAD98375 Tr:O60373 Tr:Q29280 Tr:Q86JP7 Tr:Q86QS0
                     Tr:Q8I7D3 Tr:Q8UVG1 Tr:Q90YT6 Tr:Q9TTX8"
                     /pseudo
                     /codon_start=1
                     /product="ribosomal protein L32 (RPL32) pseudogene"
     gene            complement(join(53464..53610,53925..54274))
```

Db   82162 AGCCCCACCAAGCGCGCCACCCGCCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 82221
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40

Db   82222 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGGCCCCTGAGTCCCCAACCTGCT 82281
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
Tr:AAM68299 Tr:Q8AVW0 Tr:Q8C2K0 Tr:Q90YV6 Tr:Q9NQ02

Db   82282 GCCTGCTCTGGCCACCTGGGTGTGGGAGTGGTGCCGGGTGGTGCCCTTCTGCTTCCGGCGCT 82341
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 AlaGlyIleAlaSerSerAlaValGluProValAlaProAlaAlaCysAlaCysLeu 80

Db   82342 GCCCGGGATGGCTCCAGCGCTGTGAGCCTGTGCGGGGATGCAGCCCCTGCCTGTCTA 82401
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100

Db   82402 CTGAGGACTTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAAATGGAGTGC 82461
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro 120

Db   82462 CCCCCACCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCCAGCGGCTCAAGTCAACCA 82521
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   121 TrpAlaAlaAlaSerAlaThrProMetLeuSerLysAlaSerLeuCysIleProThr 140

Db   82522 TGGGCAGCAGCTTCAGCTACCCCATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC 82581
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle 160

Db   82582 CGAGGGCCACCTCCGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGGATC 82641
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   161 ProHisProCysAspThrAlaCysSerArgSerSerSerArgSer 212

Db   82642 CCCCACCATGGACGACACAGCCTGCCCAGCACCTTTGCCAGTCCTCCGTGGCTCCGAGG 82701
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys 200

Db   82702 AGTACTATTCTTTCCATGAGTCGGACCCTGGACCCTGCAGTGGGAGATGGGCAGTGGCTCCATGT 82761

Qy   201 ArgAlaGluLysLeuMetCysSerSerSerArgSer 212
Db   82762 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 82797

RESULT 5
AL592304
LOCUS   AL592304   111738 bp   DNA   linear   HTG 25-JUL-2001
DEFINITION   Homo sapiens chromosome 1 clone RP3-426N7, 7 unordered pieces.
ACCESSION   AL592304
VERSION   AL592304.1 GI:14586390
KEYWORDS   HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE   1
AUTHORS   Mclay,K.
TITLE   Direct Submission
JOURNAL   Submitted (24-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
                CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
                requests: clonerequest@sanger.ac.uk
COMMENT   -------------- Genome Center
                Center: Sanger Centre
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquery@sanger.ac.uk
                -------------- Project Information
                Center project name: dJ426N7
                -------------- Summary Statistics
                Assembly program: XGAP4; version 4.5
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Consensus quality: 110287 bases at least Q40
                Consensus quality: 110500 bases at least Q30

CDS   /locus_tag="RP11-344H11.4-001"
                /pseudo
                complement(join(53464..53610,53925..54274))
                /locus_tag="RP11-344H11.4-001"
                /note="match: proteins: Sw:P23358 Sw:P30050 Sw:P35979
                Tr:Q862X1 Tr:Q8AVW0 Tr:Q8C2K0 Tr:Q90YV6 Tr:Q9NQ02"
                /pseudo
                /codon_start=1
                /product="ribosomal protein L12 (RPL12) pseudogene"
gene   74033..84877
                /gene="RP11-344H11.3"
                /locus_tag="RP11-344H11.3-001"
mRNA   join(74033..74089,82027..83097,83683..83757,84288..84877)
                /gene="RP11-344H11.3"
                /locus_tag="RP11-344H11.3-001"
                /product="novel protein"
                /note="match: cDNAs: Em:AK007551.1 Em:AK091952.1
                Em:AX747274.1 Em:BC006890.1 Em:BC033143.1"
misc_feature   79656..79827
                /gene="RP11-344H11.3"
                /locus_tag="RP11-344H11.3-001"
                /note="Sequence from uni-directional dGTP big dye
                terminator reads only."
CDS   join(82059..83097,83683..83757,84288..84370)
                /gene="RP11-344H11.3"
                /locus_tag="RP11-344H11.3-001"
                /note="match: proteins: Tr:Q8C1Q5 Tr:Q8N0S7 Tr:Q8NAX2"
                /standard_name="OTTHUMP00000004472"
                /codon_start=1
                /product="novel protein"
                /protein_id="CAI13559.1"
                /db_xref="GI:55959149"
                /db_xref="UniProt/TrEMBL;Q5QP32"
                /translation="MPRPGHPRPASGPPRLGPWERPTELCLETYDKPQPPPSRRTRR
                PDPKDPGHHGPESITFISGSAEPALESPTCCLLWRPWVWEWCRAAFCFRRCRDCLQRC
                GACVRGCSPCLSTBDSTEGTABEANWAKEHNGVPPSPDRAPSRRDGQRLKSTMGSSFS
                YPDVKLKGIPVYPYRATSPAPDADSCCKEPLADPPMMRHSLPSTFASSPRGSEEYYS
                FHESDLDLPEMGSGSMSSREIDVLIFKKLTELFSVHQIDELAKCTSDTVFLEKTSKIS
                DLISSITQDYHLDEQDAEGRIVRGIIRISTRKSRARPQTSEGRSTRAAAPTAAAPDSG
                HETMVGGSLSQDELTVQISQETTADAIARKLRPYGAPGYPASHDSSFQGTDTDSSGAP
                LLQVYC"
polyA_site   84877
                /gene="RP11-344H11.3"
                /locus_tag="RP11-344H11.3-001"
gene   complement(join(87577..88309,88753..88871,88966..89049,
                91374..91568,91663..91779,92621..92686,92779..92982,
                110273..110350,112510..112713))
                /gene="NUDC"
mRNA   complement(join(87577..88309,88753..88871,88966..89049,
                91374..91568,91663..91779,92621..92686,92779..92982,
                110273..110350,112510..112713))
                /gene="NUDC"
                /locus_tag="RP11-344H11.2-001"
                /product="nuclear distribution gene C homolog (A.
                nidulans)"
                /note="match: ESTs: Em:BB798032.1
                match: cDNAs: Em:AK012321.1 Em:AL136725.1 Em:BC007280.1
                Em:BC015153.1"

Alignment Scores:
Pred. No.:   9.41e-191   Length: 140207
Score: 212.00   Matches: 212
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: Gaps: 8

US-09-989-890-238 (1-212) x AL356390 (1-140207)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
                      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db   95829 AGCCCCACCAAGCCGCCGACCCGTAGACACCCAGGACCCAGGCAGCCACCATGGGC 95888
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40

Db   95889 CAGAGAGCATTACCTTCATCTCGGCTCTCGTGAGCCGGCCGCCCTTGAGTCCCCACCTGCT 95948
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60

Db   95949 GCCTGCTCTGGCGACCCCTGGGTGTGGGAGTGGTCGGCCGGCTGCCTTCTGCTTCCGCCGCT 96008
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 AlaGlyIleAlaSerAlaValGluProValCysGlyValAspAlaAlaProAlaCysLeu 80

Db   96009 GCCGGGATTGCCTCCAGCGTGTGGAGCCTGTGTGTGGGGATGCAGCCCCTGCCTGTCTA 96068
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      81 LeuArgThr-ProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCy 100

Db   96069 CTGAGGAT-TCCACTGAGGGGACTGCTGAGCCAACTGGGCCAAGGAGCACAATGGAGTG 96127
           ||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||
Qy     100 sProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPr 120

Db   96128 CCCCCCAGCCTGATCGTGCACCCCCCAGCCGCGGGGATGGCCAGCGGCTCAAGTCAACC 96187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     120 oTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProTh 140

Db   96188 ATGGGCAGCAGCTTCAGCTACACCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTAC 96247
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     140 rArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIl 160

Db   96248 CCGAGGGCCACCTCCCCAGCCCTGATGCCGGACTCCTGCTCGAAGGAGCCACTGGCCGAT 96307
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     160 eProHisProCysAspThrAlaCysSerAlaProLeuProValValLeuValAlaProAr 180

Db   96308 CCCCCACCATGGGACACAGCCTGCCCAGCACCTTTGCCAGTAGTCCTCGTGGCTCCGAG 96367
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     180 gSerThrIleLeuSerMetSerArgThrTrpThrCysSerArgArgTrpAlaValAlaProCy 200

Db   96368 GAGTACTATTCTTTCCATGAGTCCGACCTGGACCTGGCAGTGGCTCCATG 96427
           |||||||||||||||||||||||||||||||||||||||||||||||||
Qy     200 sArgAlaGluLysMetCysSerSerSerArgSer 212

Db   96428 TCGAGCCGGAGAAATTGATGTGCTCATCTTCAAGAAGC 96464
```

```
RESULT 6
AX747274
LOCUS       AX747274        1785 bp    mRNA      linear      PAT 20-JUN-2003
DEFINITION  Sequence 799 from Patent EP1308459.
ACCESSION   AX747274
VERSION     AX747274.1  GI:3213662
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
  TITLE     Full-length cDNA sequences
  JOURNAL   Patent: EP 1308459-A 799 07-MAY-2003;
            Helix Research Institute (JP); Research Association for
            Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1..1785
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
Consensus quality: 110681 bases at least Q20
Insert size: 111138; sum-of-contigs
Insert size: 119403; 8.4% error; agarose-fp
Quality coverage: 11.23x in Q20 bases; sum-of-contigs Quality
coverage: 10.67x in Q20 bases; agarose-fp
------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*          1     18058: contig of 18058 bp in length
*      18059     18158: gap of 100 bp
*      18159     35144: contig of 16986 bp in length
*      35145     35244: gap of 100 bp
*      35245     54710: contig of 19466 bp in length
*      54711     54810: gap of 100 bp
*      54811     72936: contig of 18126 bp in length
*      72937     73036: gap of 100 bp
*      73037     92888: contig of 19852 bp in length
*      92889     92988: gap of 100 bp
*      92989    108739: contig of 15751 bp in length
*     108740    108839: gap of 100 bp
*     108840    111738: contig of 2899 bp in length.
FEATURES             Location/Qualifiers
     source          1..111738
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP3-426N7"
                     /clone_lib="RPCI-3"
     misc_feature    1..18058
                     /note="assembly_fragment:02048
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    18159..35144
                     /note="assembly_fragment:02454
                     fragment_chain:1"
     misc_feature    35245..54710
                     /note="assembly_fragment:02786
                     fragment_chain:1"
     misc_feature    54811..72936
                     /note="assembly_fragment:00223
                     fragment_chain:2"
     misc_feature    73037..92888
                     /note="assembly_fragment:01820
                     fragment_chain:2"
     misc_feature    92989..108739
                     /note="assembly_fragment:01122
                     fragment_chain:2"
     misc_feature    108840..111738
                     /note="assembly_fragment:02919
                     fragment_chain:2
                     clone_end:SP6
                     vector_side:right"
ORIGIN
```

```
Alignment Scores:
Pred. No.:         1.72e-111      Length:              111738
Score:             129.00         Matches:             211
Percent Similarity: 99.1%         Conservative:        0
Best Local Similarity: 99.1%      Mismatches:          1
Query Match:       60.8%          Indels:              2
DB:                14             Gaps:                0

US-09-989-890-238 (1-212) x AL592304 (1-111738)

Qy    1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
```

```
Alignment Scores:
Pred. No.:         1.2e-99        Length:              1785
```

```
Score:                   115.00      Matches:         211
Percent Similarity:      99.1%       Conservative:      0
Best Local Similarity:   99.1%       Mismatches:        1
Query Match:             54.2%       Indels:            2
DB:                                  Gaps:              0

US-09-989-890-238 (1-212) x AX747274 (1-1785)

Qy    1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
Db  193 AGCCCCCACCAGCCGCCGCACCCGTAGACCAGACCCCCAGGACACCCTGCCACCATGGGC 252
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla     40
Db  253 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGGCCGCTCCCCCCACCTGCT     312
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   41 AlaCySerGlyAspProGlyCySGlySerGlyAlaGlyLeuProSerAlaSerAlaAla   60
Db  313 GCCTGCTCTGGGGACCCTGGGTGTGGGAGTGGTGTGCCGGGCTGCCTTCTGCTTCCGCCGCT 372
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCySGlyAspAlaAlaProAlaCysLeu  80
Db  373 GCCGGGATGGCCTCCCAGCCGTGTGGAGCCTGTGTGCGGGGATGCAGCCCCTGCCTGTCTA 432
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81 LeuArgGlyThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
Db  433 CTGAGGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGCCAAGGAGCACAATGGAGTGC 492
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  101 ProProAlaLeuIleValHisProProLeuAlaGlyMetAlaSerGly-SerSerGlnPr 120
Db  493 CCCCCAGCCCTGATCGTGCACCCCCCAGCCCGGGATGGCCAGCAG-CTCAAGTCAACC    551
        |||||||||||||||||||||||||||||||||| | |||||||| |||||||||||||
Qy  120 -TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProTh 140
Db  552 ATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGACATCCCTGTGTATCCCTAC 611
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  140 rArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIl    160
Db  612 CCGAGGGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGAT 671
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  160 eProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProAr 180
Db  672 CCCCCACCCATGCGACACAGCCTGCCCGCCCCACCTTTGCCAGTAGTCCTCGTGGCTCCGAG 731
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  180 gSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCy 200
Db  732 GAGTACTATTCTTTCCATGAGTCGGACCTGGACTGGCGCGGAGATGGGCAGTGGCTCCATG 791
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  200 sArgAlaGluLysLeuMetCysSerSerSerArgSer                       212
Db  792 TCGAGCCCAGAAATTGATGTGTCTCATCTTCAAGAAGC                      828
        ||||||||||||||||||||||||||||||||||||||
```

```
RESULT 7
LOCUS       AK091952                1785 bp    mRNA    linear    PRI 30-JAN-2004
DEFINITION  Homo sapiens cDNA FLJ34633 fis, clone KIDNE2015710.
ACCESSION   AK091952
VERSION     AK091952.1  GI:21750433
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
            Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
            Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,M.,
            Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
            Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
            Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
```

```
            Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
            Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
            Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
            Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
            Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
            Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,
            Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
            Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
            Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
            Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
            Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
            Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
            Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
            Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
            Inagaki,H., Ikena,Y., Okamoto,S., Okitani,R., Kawakami,T.,
            Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
            Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
            Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
            Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
            Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
            Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
  TITLE     Complete sequencing and characterization of 21,243 full-length
            human cDNAs
  JOURNAL   Nat. Genet. 36 (1), 40-45 (2004)
   PUBMED   14702039
REFERENCE   2
  AUTHORS   Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
            Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1785)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..1785
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="KIDNE2015710"
                     /tissue_type="kidney"
                     /clone_lib="KIDNE2"
                     /note="cloning vector: pME18SFL3"
     CDS             90..1286
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="BAC03775.1"
                     /db_xref="GI:21750434"
                     /translation="MPRPGHPRPASGPPRLGPWERPTELCLETYDKPPQPPSRRTRR
                     PDPKDPGHHGPSITFISGSAEPALESPTCCLLWRPWVWEWCRAAFCFRRCRDCLQRC
                     GACVRGCSPCLSTEDSTEGTARANWAKEHNGVPPSPDRAPPSRRDGQLKSTWGSSFS
                     YPDVKLKGIPVPYPRATSPAPDADSCCKEPLADPPMRHSLPSTFASSPRGSEBEYYS
                     FHESDLDLPEMGSGSMSSREIDVLIFKKLTELFSVHQIDELAKCTSDTVFLEKTSKIS
                     DLISSITQDYHLDEQDAEGRLVRGIIRISTRKSRARPQTSEGRSTRAAAPTAAATSIS
                     HETMVGSGLSQDELTVQISQETTADAIARKLRPYGAPGYPASHDSSFQGTDTDSSGAP
                     LLQVYC"
ORIGIN
```

US-09-989-890-238 (1-212) x AX535029 (1-1977)

```
Qy      1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357  AGCCCCCACCAAGCCAGCCGCCGCACCGTAGACCAAGGACCCCTGGCACCATGGGC 416

Qy     21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417  CAGAGAGCAGCATTACCTTCATCTCTGGCTCTGCTAGCCGCCGGCCCCTTGAGTCCCCACCCTGCT 476

Qy     41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla 59
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477  GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGCGGCTGCCTTCGCC 533
```

RESULT 10
```
AX535090
LOCUS       AX535090                1977 bp    DNA        linear    PAT 22-NOV-2002
DEFINITION  Sequence 117 from Patent WO02068633.
ACCESSION   AX535090
VERSION     AX535090.1  GI:25261789
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.
  TITLE     Compositions and methods relating to lung specific genes and
            proteins
  JOURNAL   Patent: WO 02068633-A 117 06-SEP-2002;
            Diadexus, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1977
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

Alignment Scores:
```
Pred. No.:                            Length:             1977
Score:              4.47e-46          Matches:              59
Percent Similarity: 59.00             Conservative:          0
Best Local Similarity: 100.0%         Mismatches:            0
Query Match:        100.0%            Indels:                0
DB:                 27.8%             Gaps:                  0
                      6
```

US-09-989-890-238 (1-212) x AX535090 (1-1977)

```
Qy      1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357  AGCCCCCACCAAGCCAGCCGCCGCACCGTAGACCAAGGACCCCTGGCACCATGGGC 416

Qy     21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417  CAGAGAGCAGCATTACCTTCATCTCTGGCTCTGCTAGCCGCCGGCCCCTTGAGTCCCCACCCTGCT 476

Qy     41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla 59
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477  GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGCGGCTGCCTTCGCC 533
```

RESULT 11
```
BC006890
LOCUS       BC006890                1680 bp    mRNA       linear    ROD 29-JUN-2004
DEFINITION  Mus musculus RIKEN cDNA 1810019J16 gene, mRNA (cDNA clone MGC:11921
            IMAGE:3599314), complete cds.
ACCESSION   BC006890
VERSION     BC006890.1  GI:13905189
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

(second column)

```
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1680)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1680)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 16 Row: p Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 19526881.
FEATURES             Location/Qualifiers
     source          1..1680
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:11921 IMAGE:3599314"
                     /tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
                     ductal carcinoma. 5 month old virgin mouse."
                     /clone_lib="NCI_CGAP_Mam6"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..1680
                     /gene="1810019J16Rik"
                     /db_xref="GeneID:69073"
                     /db_xref="MGI:1916323"
     CDS             142..792
                     /gene="1810019J16Rik"
                     /codon_start=1
                     /product="RIKEN cDNA 1810019J16"
                     /protein_id="AAH06890.1"
```

```
                           /db_xref="GI:13905190"
                           /db_xref="GeneID:69073"
                           /db_xref="MGI:1916323"
                           /translation="MSAPSPHRAVAPGGQTLRTLATMGQRVSPSFQALQNQPTSPQPA
                           ASSGAPGVGTGVGLPSASAAAGIACSAVGLVCGAAAPAYLPETPLKGLPKPPGPRNTM
                           VCPPARTVHPPAAGMARGSRPAWAAASATLMLSSKASLSTPTAMPPQSLTWTPAARS
                           PWPSLLPHGTACLAPSTPAPALRSTTPSMNRTWTCLRWAVAPCRAGRSTCLFSRS"
ORIGIN

Alignment Scores:
Pred. No.:          0.331              Length:        1680
Score:              12.00              Matches:       12
Percent Similarity: 100.0%             Conservative:  0
Best Local Similarity: 100.0%          Mismatches:    0
Query Match:        5.7%               Indels:        0
DB:                 9                  Gaps:          0

US-09-989-890-238 (1-212) x BC006890 (1-1680)

Qy    53 GlyLeuProSerAlaSerAlaSerAlaAlaAlaGlyIleAla 64
            |||||||||||||||||||||||||||||||||||||||
Db   310 GGGCTGCCTTCTCCTGCTTCCGCCGCTGCAGGGATTGCC 345
```

```
RESULT 12
AC004961/c
LOCUS       AC004961             68130 bp    DNA     linear   PRI 30-SEP-2000
DEFINITION  Homo sapiens PAC clone RP5-1098J4 from 7, complete sequence.
ACCESSION   AC004961
VERSION     AC004961.2  GI:5001540
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 68130)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  PUBMED    9847074
REFERENCE   2  (bases 1 to 68130)
  AUTHORS   Mead,K., Stoneking,T., Kwok,P., Kozlowicz,P. and Langston,Y.
  TITLE     The sequence of Homo sapiens PAC clone RP5-1098J4
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 68130)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 68130)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 68130)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jun 5, 1999 this sequence version replaced gi:3213036.
            --------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: saplens@watson.wustl.edu
            --------------- Summary Statistics
            Center project name: H_DJ1098J04
            ---------------

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
```

```
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7; send
            mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-5, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994).  The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:    pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP5-88608, 200 bp overlap.
            Actual start of this clone is at base position 1 of RP5-1098J4;
            actual end is at 57367 of RP5-88608.
FEATURES             Location/Qualifiers
     source          1..68130
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP5-1098J4"
                     /clone_lib="RPCI-5"
     repeat_region   1..244
                     /rpt_family="Alu"
     repeat_region   245..463
                     /rpt_family="Mariner"
     repeat_region   474..665
                     /rpt_family="L1"
     repeat_region   673..940
                     /rpt_family="L1"
     repeat_region   3719..4005
                     /rpt_family="Alu"
     repeat_region   4049..4432
                     /rpt_family="MaLR"
     misc_feature    4650..4749
                     /note="match to EST AA330816 (NID:g1983079)"
     repeat_region   4800..4833
                     /rpt_family="(CA)n"
     repeat_region   4876..4911
                     /rpt_family="(CA)n"
     repeat_region   5482..5757
                     /rpt_family="L1"
     repeat_region   5840..6055
                     /rpt_family="L1"
     repeat_region   6133..6293
                     /rpt_family="MIR"
     repeat_region   6301..6609
                     /rpt_family="Alu"
     repeat_region   6757..6809
                     /rpt_family="L1"
     repeat_region   6810..6951
```

```
repeat_region      6976..7423
                   /rpt_family="Other"
repeat_region      7446..7693
                   /rpt_family="MaLR"
repeat_region      7802..8121
                   /rpt_family="L1"
repeat_region      8428..8719
                   /rpt_family="Alu"
repeat_region      9231..9472
                   /rpt_family="Alu"
repeat_region      9474..9603
                   /rpt_family="MIR"
repeat_region      9794..9911
                   /rpt_family="MER86"
misc_feature       10566..10853
                   /note="match to EST AA330816 (NID:g1983079)"
repeat_region      11013..11126
                   /rpt_family="Alu"
repeat_region      12991..13231
                   /rpt_family="L2"
misc_feature       13813..13918
                   /rpt_family="MIR"
                   /note="similar to Mus musculus EST AA655227 (NID:g2591381)
                   vq84a11.s1"
repeat_region      14302..14592
                   /rpt_family="Alu"
repeat_region      14796..14823
                   /rpt_family="AT_rich"
repeat_region      15561..15797
                   /rpt_family="Alu"
repeat_region      16515..16535
                   /rpt_family="(GGA)n"
repeat_region      16560..16686
                   /rpt_family="AT_rich"
repeat_region      17300..17398
                   /rpt_family="Alu"
misc_feature       17302..17398
                   /note="similar to Mus musculus EST AA619820 (NID:g2523696)
                   v161c07.s1"
misc_feature       17319..17398
                   /note="similar to Mus musculus EST AA655227 (NID:g2591381)
                   vq84a11.s1"
misc_feature       17568..17800
                   /note="match to EST AI214268 (NID:g3777869) qr28a03.x1"
                   ma63a04.r1"
repeat_region      18417..18439
                   /rpt_family="AT_rich"
repeat_region      18442..18673
                   /rpt_family="Alu"
repeat_region      19015..19130
                   /rpt_family="Alu"
repeat_region      19157..19273
                   /rpt_family="L1"
repeat_region      19274..19303
                   /rpt_family="Alu"
repeat_region      19341..19756
                   /rpt_family="AT_rich"
repeat_region      20227..20260
                   /rpt_family="L1"
repeat_region      20261..20452
                   /rpt_family="(TAAA)n"
repeat_region      20728..20992
                   /rpt_family="Alu"
repeat_region      20977..20991
                   /rpt_family="MER1_type"
repeat_region      20992..21034
                   /rpt_family="L1"
repeat_region      21035..21317
                   /rpt_family="(CATA)n"
repeat_region      21318..22078
                   /rpt_family="Alu"
repeat_region      22355..22519
                   /rpt_family="L1"
                   /rpt_family="MER2_type"


repeat_region      22599..22645
                   /rpt_family="MER2_type"
misc_feature       22945..23020
                   /note="similar to Mus musculus EST AA619820 (NID:g2523696)
                   v161c07.s1"
misc_feature       22945..23020

Alignment Scores:
Pred. No.:
Score:                   71.7
Percent Similarity:      11.00
Best Local Similarity:  100.0%
Query Match:            100.0%
DB:                       5.2%
                            8

Length:        68130
Matches:          11
Conservative:      0
Mismatches:        0
Indels:            0
Gaps:              0

US-09-989-890-238 (1-212) x AC004961 (1-68130)

Qy     24 LeuProSerSerLeuAlaLeuLeuSerArgPro 34
          |||||||||||||||||||||||||||||||||
Db  42114 CTTCCATCCAGCCTCGCTCTCCTGAGCCCGCCCC 42082

RESULT 13
AC147052
LOCUS       AC147052        170236 bp    DNA       linear   PRI 03-JAN-2004
DEFINITION  Pan troglodytes BAC clone RP43-143G1 from 7, complete sequence.
ACCESSION   AC147052
VERSION     AC147052.2  GI:40018813
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 170236)
AUTHORS     Levy,A., Kozlowicz,A. and Haglund,K.
TITLE       The sequence of Pan troglodytes BAC clone RP43-143G1
JOURNAL     Unpublished (2001)
REFERENCE   2  (bases 1 to 170236)
AUTHORS     Wilson,R.
TITLE       Sequencing of Pan troglodytes
JOURNAL     Unpublished (2001)
REFERENCE   3  (bases 1 to 170236)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (04-NOV-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 170236)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (18-DEC-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 170236)
AUTHORS     Wilson,R.
TITLE       Direct Submission
JOURNAL     Submitted (03-JAN-2004) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Dec 18, 2003 this sequence version replaced gi:38154079.
COMMENT     -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: C_PT143G01
            --------------
            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
```

```
JOURNAL    Unpublished (2001)
REFERENCE  3 (bases 1 to 172424)
AUTHORS    Wilson,R.K.
TITLE      Direct Submission
JOURNAL    Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
           Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE  4 (bases 1 to 172424)
AUTHORS    Wilson,R.K.
TITLE      Direct Submission
JOURNAL    Submitted (03-SEP-2003) Genetics, Genome Sequencing Center, 4444
           Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE  5 (bases 1 to 172424)
AUTHORS    Wilson,R.
TITLE      Direct Submission
JOURNAL    Submitted (26-SEP-2003) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On Sep 3, 2003 this sequence version replaced gi:33387079.
COMMENT    -------------- Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://genome.wustl.edu
               Contact: submissions@watson.wustl.edu
           -------------- Summary Statistics
               Center project name: C_PT004K13

           --------------
           NOTICE:  This sequence may not represent the entire insert of this
           clone.  It may be shorter because we only sequence overlapping
           clone sections once, or longer because we provide a small overlap
           between neighboring data submissions.

           This sequence was finished as follows unless otherwise noted:
           all regions were double stranded, sequenced with an alternate
           chemistry, or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by sequence
           from more than one subclone; and the assembly was confirmed by
           restriction digest.

           MAPPING INFORMATION:
           Mapping information for this clone was provided by Dr. Wes Warren,
           Department of Genetics, Washington University, St. Louis MO. For
           additional information about the map position of this sequence, see
           http://genome.wustl.edu

           SOURCE INFORMATION:
           The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
           was isolated from white blood cells obtained from a male chimpanzee
           (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
           clone and detailed information can be obtained from ResGen
           (http://www.resgen.com) or Pieter de Jong and co-workers at
           http://www.bacpac.chori.org.

           NEIGHBORING SEQUENCE INFORMATION:
           This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..172424
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /map="7"
                     /clone="RP43-4K13"
                     /clone_lib="RPCI-43"
ORIGIN

Alignment Scores:
Pred. No.:              159
Score:                  11.00
Percent Similarity:     100.0%
Best Local Similarity:  100.0%
Query Match:            5.2%
DB:                     8

Length:         172424
Matches:        11
Conservative:   0
Mismatches:     0
Indels:         0
Gaps:           0
```

```
           chemistry, or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by sequence
           from more than one subclone; and the assembly was confirmed by
           restriction digest.

           MAPPING INFORMATION:
           Mapping information for this clone was provided by Dr. Wes Warren,
           Department of Genetics, Washington University, St. Louis MO. For
           additional information about the map position of this sequence, see
           http://genome.wustl.edu

           SOURCE INFORMATION:
           The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
           was isolated from white blood cells obtained from a male chimpanzee
           (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
           clone and detailed information can be obtained from ResGen
           (http://www.resgen.com) or Pieter de Jong and co-workers at
           http://www.bacpac.chori.org.

           NEIGHBORING SEQUENCE INFORMATION:
           This sequence is the entire insert of the clone. This clone is
           overlapped by AC146125.
FEATURES             Location/Qualifiers
     source          1..170236
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /map="7"
                     /clone="RP43-143G1"
                     /clone_lib="RPCI-43"
     misc_feature    1..40
                     /note="Sequence derived from PCR product of project DNA."
     unsure          3612..32837
                     /note="Sequence derived from one plasmid subclone."
     misc_feature    33048..33060
                     /note="Sequence derived from PCR product of project DNA."
ORIGIN

Alignment Scores:
Pred. No.:              157
Score:                  11.00
Percent Similarity:     100.0%
Best Local Similarity:  100.0%
Query Match:            5.2%
DB:                     8

Length:         170236
Matches:        11
Conservative:   0
Mismatches:     0
Indels:         0
Gaps:           0

US-09-989-890-238 (1-212) x AC147052 (1-170236)

Qy     24 LeuProSerLeuAlaLeuLeuSerArgPro 34
          ||||||||||||||||||||||||||||||||||||
Db 112612 CTTCCATCCAGCCTCCGCTCCTCCTGAGCCGCCCCC 112644

RESULT 14
AC146125
LOCUS       AC146125     172424 bp    DNA     linear   PRI 26-SEP-2003
DEFINITION  Pan troglodytes BAC clone RP43-4K13 from 7, complete sequence.
ACCESSION   AC146125
VERSION     AC146125.2  GI:34419744
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1 (bases 1 to 172424)
AUTHORS     Griffin,M., Haglund,K. and Haakenson,W.
TITLE       The sequence of Pan troglodytes BAC clone RP43-4K13
JOURNAL     Unpublished (2001)
REFERENCE   2 (bases 1 to 172424)
AUTHORS     Sulston,J.E. and Wilson,R.
TITLE       Sequencing of Pan troglodytes
```

US-09-989-890-238 (1-212) x AC146125 (1-172424)

Qy   24 LeuProSerSerLeuAlaLeuLeuSerArgPro 34
                |||||||||||||||||||||||||||||||||
Db   72277 CTTCCATCCAGCTCCGCTCTCCTGAGCCGCCCC 72309

RESULT 15
AF175677/c
LOCUS   AF175677   525 bp   DNA   linear   BCT 06-SEP-1999
DEFINITION Neisseria meningitidis strain BZ232 surface protein A (nspA) gene,
            complete cds.
ACCESSION   AF175677
VERSION   AF175677.1 GI:5825528
KEYWORDS   .
SOURCE   Neisseria meningitidis
  ORGANISM Neisseria meningitidis
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1 (bases 1 to 525)
  AUTHORS   Moe,G.R., Tan,S. and Granoff,D.M.
  TITLE   Differences in Surface Expression of Neisserial Surface Protein A
            among Neisseria meningitidis Group B strains
  JOURNAL   Infect. Immun. (1999) In press
REFERENCE   2 (bases 1 to 525)
  AUTHORS   Moe,G.R., Tan,S. and Granoff,D.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-AUG-1999) Children's Hospital Oakland Research
            Institute, 747 Fifty-Second Street, Oakland, CA 94609, USA
FEATURES   Location/Qualifiers
     source   1..525
                     /organism="Neisseria meningitidis"
                     /mol_type="genomic DNA"
                     /strain="BZ232"
                     /db_xref="taxon:487"
     gene   1..525
                     /gene="nspA"
     CDS   1..525
                     /gene="nspA"
                     /codon_start=1
                     /transl_table=11
                     /product="surface protein A"
                     /protein_id="AAD53280.1"
                     /db_xref="GI:5825529"
                     /translation="MKKALATLIALAIPAAALAEGASGFYVQADAAHAKASSLGSAK
                     GFSPRISAGYRINDLRFAVDYTRYKNYKAPSTDFKLYSIGASAIYDFDTQSPVKPYLG
                     ARLSLNRASVDLGGSDSFSQTSTGLGVLTGVSYAVTPNVDLDAGYRYNYIGKVNTVKN
                     VRSGELSAGVRVKF"

ORIGIN

Alignment Scores:
Pred. No.:   Length: 525
Score: 9.97   Matches: 10
Percent Similarity: 10.00%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 4.7%   Gaps: 0
                   1

US-09-989-890-238 (1-212) x AF175677 (1-525)

Qy   55 ProSerAlaSerAlaAlaAlaGlyIleAla 64
                ||||||||||||||||||||||||||||||
Db   62 CCTTCCGCCAGTGCGGCGGCCGCCGGATAGCG 33

RESULT 16
AF175679/c
LOCUS   AF175679   525 bp   DNA   linear   BCT 06-SEP-1999
DEFINITION Neisseria meningitidis strain M136 surface protein A (nspA) gene,
            complete cds.
ACCESSION   AF175679
VERSION   AF175679.1 GI:5825533
KEYWORDS   .
SOURCE   Neisseria meningitidis

ORGANISM   Neisseria meningitidis
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1 (bases 1 to 525)
  AUTHORS   Moe,G.R., Tan,S. and Granoff,D.M.
  TITLE   Differences in Surface Expression of Neisserial Surface Protein A
            among Neisseria meningitidis Group B strains
  JOURNAL   Infect. Immun. (1999) In press
REFERENCE   2 (bases 1 to 525)
  AUTHORS   Moe,G.R., Tan,S. and Granoff,D.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-AUG-1999) Children's Hospital Oakland Research
            Institute, 747 Fifty-Second Street, Oakland, CA 94609, USA
FEATURES   Location/Qualifiers
     source   1..525
                     /organism="Neisseria meningitidis"
                     /mol_type="genomic DNA"
                     /strain="M136"
                     /db_xref="taxon:487"
     gene   1..525
                     /gene="nspA"
     CDS   1..525
                     /gene="nspA"
                     /codon_start=1
                     /transl_table=11
                     /product="surface protein A"
                     /protein_id="AAD53282.1"
                     /db_xref="GI:5825534"
                     /translation="MKKALATLIALAIPAAALAEGASGFYVQADAAHAKASSLGSAK
                     GFSPRISAGYRINDLRFAVDYTRYKNYKAPSTDFKLYSIGASAIYDFDTQSPVKPYLG
                     ARLSLNRASVDLGGSDSFSQTSTGLGVLTGVSYAVTPNVDLDAGYRYNYIGKVNTVKN
                     VRSGELSAGVRVKF"

ORIGIN

Alignment Scores:
Pred. No.:   Length: 525
Score: 9.97   Matches: 10
Percent Similarity: 10.00%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 4.7%   Gaps: 0
                   1

US-09-989-890-238 (1-212) x AF175679 (1-525)

Qy   55 ProSerAlaSerAlaAlaAlaGlyIleAla 64
                ||||||||||||||||||||||||||||||
Db   62 CCTTCCGCCAGTGCGGCGGCCGCCGGATAGCG 33

RESULT 17
AF175681/c
LOCUS   AF175681   525 bp   DNA   linear   BCT 06-SEP-1999
DEFINITION Neisseria meningitidis strain NG3/88 surface protein A (nspA) gene,
            complete cds.
ACCESSION   AF175681
VERSION   AF175681.1 GI:5825538
KEYWORDS   .
SOURCE   Neisseria meningitidis
  ORGANISM Neisseria meningitidis
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1 (bases 1 to 525)
  AUTHORS   Moe,G.R., Tan,S. and Granoff,D.M.
  TITLE   Differences in Surface Expression of Neisserial Surface Protein A
            among Neisseria meningitidis Group B strains
  JOURNAL   Infect. Immun. (1999) In press
REFERENCE   2 (bases 1 to 525)
  AUTHORS   Moe,G.R., Tan,S. and Granoff,D.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-AUG-1999) Children's Hospital Oakland Research
            Institute, 747 Fifty-Second Street, Oakland, CA 94609, USA
FEATURES   Location/Qualifiers
     source   1..525
                     /organism="Neisseria meningitidis"

```
                                    /mol_type="genomic DNA"
                                    /strain="NG3/88"
                                    /db_xref="taxon:487"
gene            1..525
                                    /gene="nspA"
CDS             1..525
                                    /gene="nspA"
                                    /codon_start=1
                                    /transl_table=11
                                    /product="surface protein A"
                                    /protein_id="AAD53284.1"
                                    /db_xref="GI:5825539"
                                    /translation="MKKALATLIALAIPAAALAEGASGFYVQADAHAKASSSLGSAK
                                    GFSPRISAGYRINDLRPAVDYTRYKNYKAPSTDFKLYSIGASAIYDFDTQSPVKPYLG
                                    ARLSLNRASVDLGGGSDSFSQTSTGLGVLAGVSYAVTPNVDLDAGYRYNYIGKVNTVRN
                                    VRSGELSAGVRVKF"

ORIGIN

Alignment Scores:
Pred. No.:              9.97              Length:            525
Score:                 10.00%            Matches:            10
Percent Similarity:    100.0%            Conservative:        0
Best Local Similarity: 100.0%            Mismatches:          0
Query Match:           4.7%              Indels:              0
DB:                    1                 Gaps:                0

US-09-989-890-238 (1-212) x AF175681 (1-525)

Qy    55  ProSerAlaSerAlaAlaGlyIleAla 64
          |||||||||||||||||||||||||||
Db    62  CCTTCGCCAGTGCGGCGCCGGATAGCG 33


RESULT 18
STMNUSG/c
LOCUS       STMNUSG            1166 bp    DNA       linear   BCT 04-FEB-1999
DEFINITION  S.coelicolor gene for NusG, complete cds.
ACCESSION   D17465
VERSION     D17465.1  GI:435785
KEYWORDS    NusG; transcription antitermination.
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 1166)
  AUTHORS   Miyake,K., Onaka,H., Horinouchi,S. and Beppu,T.
  TITLE     Organization and nucleotide sequence of the secE-nusG region of
            Streptomyces griseus
  JOURNAL   Biochim. Biophys. Acta 1217 (1), 97-100 (1994)
   PUBMED   8286423
REFERENCE   2  (bases 1 to 1166)
  AUTHORS   Onaka,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1993) Hiroyasu Onaka, The University of Tokyo,
            Department of Biotechnology; Yayoi, Bunkyou-ku, Tokyo 113, Japan
            (E-mail:aa57093@hongo.ecc.u-tokyo.ac.jp, Tel:03-3812-2111(ex.5147),
            Fax:03-5802-2931)
COMMENT     Submitted (24-Aug-1993) to DDBJ by:
            Hiroyasu Onaka
            Dept. of Agricultural Chemistry
            Faculty of Agriculture
            The University of Tokyo
            Bunkyo-ku, Tokyo 113
            Japan
            Phone:  03-3812-2111 x5147
            Fax:    03-3812-0544.
FEATURES             Location/Qualifiers
     source          1..1166
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /sub_strain="M130"
                     /db_xref="taxon:100226"
     gene            60..962
```

```
CDS                                 /gene="nusG"
                60..962
                                    /gene="nusG"
                                    /codon_start=1
                                    /transl_table=11
                                    /product="NusG"
                                    /protein_id="BAA04282.1"
                                    /db_xref="GI:483834"
                                    /translation="MSDPNVNDAIEPVESVEBDLGTVEGADNEDTEASAEABAADDTV
                                    VAETDEDATDARADETRAEATDARADRDQTRRKAEBSEEEBAEEPELDPIEKL
                                    RPGVRVLPGEWVYIHTYAGYENRVKTNLEQRAVSLNVEDYIFQAEVPQKEVVQIKNGD
                                    RKTIRQNKLPGYVLVRMDYTNESWGVVRNTPGVTGFVGNAYDPYPLTLDEIVKMLAPE
                                    AEEKAARRAAEBGKPAPQRKVEVQVLDPBVGDSVTVTDGPFATLQATINBINADSKK
                                    VKGLVEIFGRETPVELSFDQIQKN"

ORIGIN

Alignment Scores:
Pred. No.:             19.8             Length:           1166
Score:                 10.00            Matches:            10
Percent Similarity:    100.0%           Conservative:        0
Best Local Similarity: 100.0%           Mismatches:          0
Query Match:           4.7%             Indels:              0
DB:                    1                Gaps:                0

US-09-989-890-238 (1-212) x STMNUSG (1-1166)

Qy    51  GlyAlaGlyLeuProSerAlaSerAlaAla 60
          ||||||||||||||||||||||||||||||
Db   766  GGCGCCGGCTTGCCCTCGGCCTCGGCGGCC 737


RESULT 19
BC048207/c
LOCUS       BC048207            1487 bp    mRNA      linear   PRI 20-AUG-2003
DEFINITION  Homo sapiens hypothetical protein LOC348645, mRNA (cDNA clone
            IMAGE:5419610), partial cds.
ACCESSION   BC048207
VERSION     BC048207.1  GI:29387213
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1487)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1487)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
```

REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 44 Row: h Column: 2.
FEATURES   Location/Qualifiers
     source   1..1487
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5419610"
                     /tissue_type="Lymph, Burkitt lymphoma"
                     /clone_lib="NIH_MGC_8"
                     /lab_host="DH10B-R"
                     /note="vector: pOTB7"
ORIGIN

Alignment Scores:
Pred. No.:   Length: 1487
Score: 24.4   Matches: 10
Percent Similarity: 10.00%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 4.7%   Gaps: 0
                            8

US-09-989-890-238 (1-212) x BC048207 (1-1487)

Qy   26 SerSerLeuAlaLeuSerArgProLeu 35
                |||||||||||||||||||||||||||
Db   415 TCCTCACTGGCCCTTCTATCCAGGCCCTT 386

RESULT 20
CQ841901/c
LOCUS   CQ841901   1635 bp   DNA   linear   PAT 02-AUG-2004
DEFINITION Sequence 548 from Patent EP1440981.
ACCESSION   CQ841901
VERSION   CQ841901.1 GI:50893688
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J., Isono,Y., Nagai,K. and Irie,R.
  TITLE   Full-length human cdna
  JOURNAL   Patent: EP 1440981-A 548 28-JUL-2004;
            Research Association for Biotechnology (JP)
FEATURES   Location/Qualifiers
     source   1..1635
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:   Length: 1635
Score: 26.4   Matches: 10
Percent Similarity: 10.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 4.7%   Gaps: 0
                            6

US-09-989-890-238 (1-212) x CQ841901 (1-1635)

Qy   26 SerSerLeuAlaLeuSerArgProLeu 35
                |||||||||||||||||||||||||||
Db   1332 TCCTCACTGGCCCTTCTATCCAGGCCCTT 1303

RESULT 21
AK124962/c
LOCUS   AK124962   1635 bp   mRNA   linear   PRI 09-SEP-2003
DEFINITION Homo sapiens cDNA FLJ42972 fis, clone BRSTN2019129.
ACCESSION   AK124962
VERSION   AK124962.1 GI:34530900
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,H., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
            and Isogai,T.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1635)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES   Location/Qualifiers
     source   1..1635
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="BRSTN2019129"
                     /tissue_type="subthalamic nucleus"
                     /clone_lib="BRSTN2"
                     /note="cloning vector: pME18SFL3"
     CDS   1004..1459
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC86007.1"
                     /db_xref="GI:34530901"
                     /translation="MCVTVDVEGPDDVGETLSDAVRDGLGTILRGGAEGSYDNWPHTR
                     KSWGPLSPGHQRBLWTQPDPWTEVLSGHKGDAGACGCCFCSQFINARCAHPLCLARG
                     LDRRASEMPILQALCLLPKVSTRSITVPSPQRSAPRASLCPPHKGKSP"
ORIGIN

Alignment Scores:
Pred. No.:   Length: 1635
Score: 26.4   Matches: 10
                            10.00

```
Percent Similarity:      100.0%        Conservative:    0
Best Local Similarity:   100.0%        Mismatches:      0
Query Match:               4.7%        Indels:          0
DB:                           8        Gaps:            0

US-09-989-890-238 (1-212) x AK124962 (1-1635)

Qy     26  SerSerLeuAlaLeuSerArgProLeu 35
           ||||||||||||||||||||||||||||
Db   1332  TCCTCACTGGCCCTTCTATCCAGGCCCCTT 1303

RESULT 22
BC079148
LOCUS        BC079148             1718 bp    mRNA     linear   ROD 15-SEP-2004
DEFINITION   Rattus norvegicus similar to hypothetical protein FLJ34633, mRNA
             (cDNA clone MGC:94165 IMAGE:7128510), complete cds.
ACCESSION    BC079148
VERSION      BC079148.1  GI:50927712
KEYWORDS     MGC.
SOURCE       Rattus norvegicus (Norway rat)
ORGANISM     Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE    1  (bases 1 to 1718)
AUTHORS      Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
             Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
             Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
             Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
             Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
             Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
             Scheetz,T.B., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
             Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
             Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
             McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
             Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
             Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
             Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
             Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
             Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
             Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
             Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
             Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE        Generation and initial analysis of more than 15,000 full-length
             human and mouse cDNA sequences
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED       12477932
REFERENCE    2  (bases 1 to 1718)
AUTHORS      Director MGC Project.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-2004) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
REMARK       NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Howard Jacobs
             cDNA Library Preparation: Express Genomics
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Sequencing Group at the Stanford Human Genome
             Center, Stanford University School of Medicine, Stanford, CA 94305
             Web site:    http://www-shgc.stanford.edu
             Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
             Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
             R. M.
             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAK Plate: 184 Row: d Column: 16
             This clone was selected for full length sequencing because it
             passed the following selection criteria: Hexamer frequency ORF
             analysis, Similarity but not identity to protein.
```

```
FEATURES             Location/Qualifiers
     source          1..1718
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="MGC:94165 IMAGE:7128510"
                     /tissue_type="Kidney, rat (Brown Norway)"
                     /clone_lib="NIH_MGC_236"
                     /lab_host="DH10B"
                     /note="Vector: pExpress1"
     gene            1..1718
                     /gene="MGC94165"
                     /db_xref="GeneID:313018"
     CDS             73..1266
                     /gene="MGC94165"
                     /codon_start=1
                     /product="similar to hypothetical protein FLJ34633"
                     /protein_id="AAH79148.1"
                     /db_xref="GI:50927713"
                     /db_xref="GeneID:313018"
                     /translation="MPRPGQPRPSSGGPPRLGPWERPTELCLETNDERSQPPGRRTRR
                     PDPKDPGHHGPRSITFISGSAEPANEPTCCLLWRPWGWDWCRAAFCFRRCRDCLQRC
                     GACVRSCSPCLSAGDPIEGSSEAAWAKEHNGVPPSPDRAPPSRRDGQKLKTSMGSSFS
                     YPDVLKGIPVYPVRHATSPVPDADSCCKEPLADPPTRHSLPSTFTSSPRGSBEYYS
                     FHESDLDLPEMGSGSMSSREIDVLIFKKLTELPSVHQIDELAKCTSDVVFLEKTSKIS
                     DLISSITQDYHLDEQDAEGRLVRGIIRISTRKSRSRPQTSEGRSARSTAPAAAPDSGH
                     ETMVGSGLSQDELTVQISQETTADAIARKLRPYGAPGYPASQDSSFQGTDTDSSGAPL
                     LQYYC"

ORIGIN

Alignment Scores:
Pred. No.:                  27.6          Length:              1718
Score:                      10.00         Matches:               10
Percent Similarity:        100.0%         Conservative:           0
Best Local Similarity:     100.0%         Mismatches:             0
Query Match:                 4.7%         Indels:                 0
DB:                            9          Gaps:                   0

US-09-989-890-238 (1-212) x BC079148 (1-1718)

Qy     60  AlaAlaAlaGlyIleAlaSerSerAlaValGlu 69
           |||||||||||||||||||||||||||||||||
Db    353  GCTGCAGGAGGATTGCCTCCAGCGCTGTGGAG 382

RESULT 23
BT007455
LOCUS        BT007455             1968 bp    mRNA      linear   PRI 13-MAY-2003
DEFINITION   Homo sapiens forkhead box O1A (rhabdomyosarcoma) mRNA, complete
             cds.
ACCESSION    BT007455
VERSION      BT007455.1  GI:30583748
KEYWORDS     FLI_CDNA.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE    1  (bases 1 to 1968)
AUTHORS      Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
             Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
             Phelan,M. and Farmer,A.
TITLE        Cloning of human full-length CDSs in BD Creator(TM) System Donor
             vector
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 1968)
AUTHORS      Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
             Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
             Phelan,M. and Farmer,A.
TITLE        Direct Submission
JOURNAL      Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
             Circle, Palo Alto, CA 94303, USA
COMMENT      This CDS clone is a part of a collection of human full length
             expression clones generated by BD Biosciences Clontech and the
```

Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: http://bioinfo.clontech.com/orfclones.

```
FEATURES             Location/Qualifiers
     source          1..1968
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GH00949X1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1..1968
                     /codon_start=1
                     /product="forkhead box O1A (rhabdomyosarcoma)"
                     /protein_id="AAP36123.1"
                     /db_xref="GI:30583749"
                     /translation="MAEAPQVVEIDPDFEPLPRPRSCTWPLPRPEFSQSNSATSSPAP
                     SGSAAANPDAAAGLPSASAAAVSADFMSNLSLLEESEDFPQAPGSVAAAVAAAAAAAA
                     TGGLCGDFQGPEAGCLHPAPPQPPPPGPLSQHPPVPAAAGPLAGQPRKSSSSRRNAW
                     GNLSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLS
                     LHSKFIRVQNEGTGKSSWWMLNPEGGKSGKSPRRRAASMDNNSKFAKSRSRAAKKKAS
                     LQSGQEGAGDSPGSQFSKWPASPGSHSNDDFDNWSTFRPRTSSNASTISGRLSPIMTE
                     QDDLGEGDVHSMVYPPSAAKMASTLPSLSEISNPENMENLLDNLNLLSSPTSLTVSTQ
                     SSPGTMWQQTPCYSFAPPNTSLNSPSPNYQKYTYGQSSMSPLPQMPIQTLQDNKSSYG
                     GMSQYNCAPGLLKELLTSDSPPHNDIMTPVDPGVAQPNSRVLGQNVMMGPNSVMSTYG
                     SQASHNKMMNPSSHTHPGHAQQTSAVNGRPLPHTVSTMPHTSGMNRLTQVKTPVQVPL
                     PHPMQMSALGGYSSVSSCNGYGRMGLLHQEKLPSDLDGMFIERLDCDMESIIRNDLMD
                     GDTLDFNFDNVLPNQSFPHSVKTTTHSWVSG"
ORIGIN

Alignment Scores:
Pred. No.:              31            Length:        1968
Score:                  10.00        Matches:       10
Percent Similarity:     100.0%       Conservative:  0
Best Local Similarity:  100.0%       Mismatches:    0
Query Match:            4.7%          Indels:        0
DB:                     8            Gaps:          0

US-09-989-890-238 (1-212) x BT007455 (1-1968)

Qy    52 AlaGlyProSerAlaSerAlaAlaAla 61
         |||||||||||||||||||||
Db   166 GCGGGCCCTGCCCCTCGGCCCTGCCCGCT 195

RESULT 24
AY890173
LOCUS       AY890173    1968 bp    mRNA    linear    SYN 29-MAR-2005
DEFINITION  Synthetic construct Homo sapiens clone FLH015808.01X forkhead box
            O1A (FOXO1A) mRNA, complete cds.
ACCESSION   AY890173
VERSION     AY890173.1  GI:61361857
KEYWORDS    Human ORF Project.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1968)
  AUTHORS   Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
            Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
            Williamson,J. and LaBaer,J.
  TITLE     Cloning of human full-length CDS in Creator (TM) recombinational
            vector system
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1968)
  AUTHORS   Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
            Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
            Williamson,J. and LaBaer,J.
```

```
TITLE     Direct Submission
JOURNAL   Submitted (04-JAN-2005) Biological Chemistry and Molecular
          Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
          Cambridge, MA 02141, USA
COMMENT   This CDS clone is a part of a collection of human full-length
          expression clones generated by Harvard Institute of Proteomics. This
          ORF clone has been cloned using normalized stop-codon. The CDS
          has been directionally cloned using BD In-Fusion(TM) cloning system
          between the SalI and HindIII sites of the pDNR-Dual vector.
          Additional sequences in the clone: 'ACC' after SalI site and
          before 'ATG' to provide Kozak consensus sequence. Each clone is
          clonally isolated and full-length sequence-verified.
FEATURES             Location/Qualifiers
     source          1..1968
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="FLH015808.01X"
                     /lab_host="Escherichia coli DH5alpha T1 resistant"
                     /note="derived from MGC template"
     gene            1..1968
                     /gene="FOXO1A"
     CDS             1..1968
                     /gene="FOXO1A"
                     /note="rhabdomyosarcoma"
                     /codon_start=1
                     /transl_table=11
                     /product="forkhead box O1A"
                     /protein_id="AAX42115.1"
                     /db_xref="GI:61361858"
                     /translation="MAEAPQVVEIDPDFEPLPRPRSCTWPLPRPEFSQSNSATSSPAP
                     SGSAAANPDAAAGLPSASAAAVSADFMSNLSLLEESEDFPQAPGSVAAAVAAAAAAAA
                     TGGLCGDFQGPEAGCLHPAPPQPPPPGPLSQHPPVPAAAGPLAGQPRKSSSSRRNAW
                     GNLSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLS
                     LHSKFIRVQNEGTGKSSWWMLNPEGGKSGKSPRRRAASMDNNSKFAKSRSRAAKKKAS
                     LQSGQEGAGDSPGSQFSKWPASPGSHSNDDFDNWSTFRPRTSSNASTISGRLSPIMTE
                     QDDLGEGDVHSMVYPPSAAKMASTLPSLSEISNPENMENLLDNLNLLSSPTSLTVSTQ
                     SSPGTMWQQTPCYSFAPPNTSLNSPSPNYQKYTYGQSSMSPLPQMPIQTLQDNKSSYG
                     GMSQYNCAPGLLKELLTSDSPPHNDIMTPVDPGVAQPNSRVLGQNVMMGPNSVMSTYG
                     SQASHNKMMNPSSHTHPGHAQQTSAVNGRPLPHTVSTMPHTSGMNRLTQVKTPVQVPL
                     PHPMQMSALGGYSSVSSCNGYGRMGLLHQEKLPSDLDGMFIERLDCDMESIIRNDLMD
                     GDTLDFNFDNVLPNQSFPHSVKTTTHSWVSG"
ORIGIN

Alignment Scores:
Pred. No.:              31            Length:        1968
Score:                  10.00        Matches:       10
Percent Similarity:     100.0%       Conservative:  0
Best Local Similarity:  100.0%       Mismatches:    0
Query Match:            4.7%          Indels:        0
DB:                     11           Gaps:          0

US-09-989-890-238 (1-212) x AY890173 (1-1968)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||
Db   166 GCGGGCCCTGCCCCTCGGCCCTGCCCGCT 195

RESULT 25
AY890174
LOCUS       AY890174    1968 bp    mRNA    linear    SYN 29-MAR-2005
DEFINITION  Synthetic construct Homo sapiens clone FLH015809.01X forkhead box
            O1A (FOXO1A) mRNA, complete cds.
ACCESSION   AY890174
VERSION     AY890174.1  GI:61361865
KEYWORDS    Human ORF Project.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1968)
  AUTHORS   Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
            Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
            Williamson,J. and LaBaer,J.
```

```
TITLE       Cloning of human full-length CDS in Creator (TM) recombinational
            vector system
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 1968)
AUTHORS     Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
            Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,B.,
            Williamson,J. and LaBaer,J.
TITLE       Direct Submission
JOURNAL     Submitted (04-JAN-2005) Biological Chemistry and Molecular
            Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
            Cambridge, MA 02141, USA
COMMENT     This CDS clone is a part of a collection of human full-length
            expression clones generated by Harvard Institute of Proteomics.
            This ORF clone has been cloned with normalized stop-codon. The CDS
            has been directionally cloned using BD In-Fusion(TM) cloning system
            between the SalI and HindIII sites of the pDNR-Dual vector.
            Additional sequences in the clone: 'ACC' after SalI site and
            before 'ATG' to provide Kozak consensus sequence. Each clone is
            clonally isolated and full-length sequence-verified.
FEATURES    Location/Qualifiers
     source 1..1968
            /organism="synthetic construct"
            /mol_type="mRNA"
            /db_xref="taxon:32630"
            /clone="FLH015809.01X"
            /lab_host="Escherichia coli DH5alpha T1 resistant"
            /note="derived from MGC template"
     gene   1..1968
            /gene="FOXO1A"
     CDS    1..1968
            /gene="FOXO1A"
            /note="rhabdomyosarcoma"
            /codon_start=1
            /transl_table=11
            /product="forkhead box O1A"
            /protein_id="AAX42116.1"
            /db_xref="GI:61361866"
            /translation="MAERAPQVVEIDPDFEPLRPRSCTWPLPRPEFSQSNSATSSPAP
            SGSAAANPDAAAGLPSASAAAVSADFMNLSLLEESEDFPQAPGSVAAAVAAAAAAAA
            TGGLCGDFQGPRAGCLHPAPPQPPPPGPLSQHPPVPAAAGPLAGQPRKSSSSRRNAW
            GNLSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLS
            LHSKFIRVQNEGTGKSSWWMLNPEGGKSGKSPRRAASMDNNSKFAKSRSRAAKKKAS
            LQSGQEGAGDSPGSQFSKWPASPGSHSNDDFPDNWSTFRPRTSSNASTISGRLSPIMTE
            QDDLGEGDVHSMVYPPSAAKMASTLPSLSEISNPENMENLLDNLNLLSSPTSLTVSTQ
            SSPGTMMQQTPCYSFAPPNTSLNSPSPNYQKYTYGQSSMSPLPQMPIQTLQDNKSSYG
            GMSQYNCAPGLLKELLTSDSPPHNDIMTPVDPGVAQPNSRVLGQNVMWGPNSVMSTYG
            PHPMQMSALGGYYSVSSCNGYGRMGLLHQEKLPSDLDGMFIERLDCMESIIRNDLMD
            GDTLDFNFDNVLPNQSPFHSVKTTTHSWVSG"
ORIGIN

Alignment Scores:
Pred. No.:                 31             Length:              1968
Score:                     10.00          Matches:             10
Percent Similarity:        100.0%         Conservative:        0
Best Local Similarity:     100.0%         Mismatches:          0
Query Match:               4.7%           Indels:              0
DB:                        11             Gaps:                0

US-09-989-890-238 (1-1968) x AY890174 (1-1968)

Qy    52 AlaGlyProSerAlaSerAlaAlaAla 61
         ||| ||| ||| ||| ||| ||| |||
Db   166 GCGGGCCTGCCCTCGGCTCGGCTGCCGCT 195

RESULT 26
LOCUS       STMSEC          1975 bp    DNA    linear   BCT 04-FEB-1999
DEFINITION  S.griseus genes for SecB and NusG, complete cds.
ACCESSION   D17464
VERSION     D17464.1  GI:436786
KEYWORDS    NusG; SecB; transcription antitermination.
SOURCE      Streptomyces griseus
```

```
ORGANISM    Streptomyces griseus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 1975)
AUTHORS     Miyake,K., Onaka,H., Horinouchi,S. and Beppu,T.
TITLE       Organization and nucleotide sequence of the secB-nusG region of
            Streptomyces griseus
JOURNAL     Biochim. Biophys. Acta 1217 (1), 97-100 (1994)
PUBMED      8286423
REFERENCE   2  (bases 1 to 1975)
AUTHORS     Onaka,H.
TITLE       Direct Submission
JOURNAL     Submitted (24-AUG-1993) Hiroyasu Onaka, The University of Tokyo,
            Department of Biotechnology; Yayoi, Bunkyou-ku, Tokyo 113, Japan
            (E-mail:aa57093@hongo.ecc.u-tokyo.ac.jp, Tel:03-3812-2111(ex.5147),
            Fax:03-5802-2931)
COMMENT     Submitted (24-Aug-1993) to DDBJ by:
            Hiroyasu Onaka
            Dept. of Agriculture Chemistry
            Faculty of Agriculture
            The University of Tokyo
            Bunkyo-ku, Tokyo 113
            Japan
            Phone: 03-3812-2111 x5147
            Fax:   03-3812-0544.
FEATURES    Location/Qualifiers
     source 1..1975
            /organism="Streptomyces griseus"
            /mol_type="genomic DNA"
            /strain="IFO1350"
            /db_xref="taxon:1911"
     gene   612..872
            /gene="secB"
     CDS    612..872
            /gene="secB"
            /codon_start=1
            /transl_table=11
            /product="SecB"
            /protein_id="BAA04280.1"
            /db_xref="GI:483835"
            /translation="MPDADEAPBSKKKSRKGGKRGKGGPLGRLALFYRQIVABLRKV
            VWPTRSQLTTYTSVVIFVVVMIGLVTVLDMGFARVVKVYFG"
     gene   946..1830
            /gene="nusG"
     CDS    946..1830
            /gene="nusG"
            /codon_start=1
            /transl_table=11
            /product="NusG"
            /protein_id="BAA04281.1"
            /db_xref="GI:483836"
            /translation="MSDPNLNDAVEPRAGASEAKDELDIVEAADSVDPDQABABDLA
            AGEPAERAAVNVAGDDSDEDDAAABEAVEADDESADEEBRAGRRPVDPVAALRDELRT
            LPGEWYIHTYAGYEKRVKANLEQRAVSLNVEEFIYQAEVPBEEIVQIKNGERKNVRQ
            NKLPGYVLVRMDLTNESWGVVRNTPGVTGFVGNAYDPYPLTLDEIVKMLAPEABEKAA
            REAABAEGKPAPARKVEVQVLDFBVGDSVTVTDGPPATLQATINEINADSKKVKGLVE
            IFGRETPVDVSFDQIQKN"
ORIGIN

Alignment Scores:
Pred. No.:                 31.1           Length:              1975
Score:                     10.00          Matches:             10
Percent Similarity:        100.0%         Conservative:        0
Best Local Similarity:     100.0%         Mismatches:          0
Query Match:               4.7%           Indels:              0
DB:                        1              Gaps:                0

US-09-989-890-238 (1-212) x STMSEC (1-1975)

Qy    51 GlyAlaGlyLeuProSerAlaSerAlaAla 60
         ||||||||||||||||||||||||||||||
Db  1634 GGAGCGGGCTTGCCCTCGGCCTCGGCGGCC 1605
```

```
RESULT 27
AY094061
LOCUS       AY094061              2162 bp    mRNA    linear   MAM 01-MAY-2002
DEFINITION  Sus scrofa forkhead/winged helix transcription factor FOXO1a
            (foxo1a) mRNA, complete cds.
ACCESSION   AY094061
VERSION     AY094061.1  GI:20384962
KEYWORDS    .
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
            Sus.
REFERENCE   1  (bases 1 to 2162)
  AUTHORS   Zhu,Q., Cunningham,M.A. and Hammond,J.M.
  TITLE     FKHR Expression in Porcine Granulosa Cells
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2162)
  AUTHORS   Zhu,Q., Cunningham,M.A. and Hammond,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2002) Dept. of Medicine/Div. of Endocrinology,
            Pennsylvania State Univ. College of Medicine, 500 University Dr.,
            Hershey, PA 17033, USA
FEATURES             Location/Qualifiers
     source          1..2162
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
     gene            1..2162
                     /gene="foxo1a"
     CDS             33..2021
                     /gene="foxo1a"
                     /codon_start=1
                     /product="forkhead/winged helix transcription factor
                     FOXO1a"
                     /protein_id="AAM19156.1"
                     /db_xref="GI:20384963"
                     /translation="MAEAPQVEIDPDFEPLPRPRSCTWPLPRPEFSQSNSATSSPAP
                     SGGPAANPDAAAGLPSASAAVNADFMSNLSLLEESGDFQQAPGSVAAAAAAAVAA
                     AAAAATGGLCGLCGDPGPEAGCLHPAPPQQPPPGPLSQHPPVPPAAAGSLAGQPRKSS
                     SRRNAWGNLSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKN
                     SIRHNLSLHSKFIRVQNEGTGKSSWWMLNPEGGKSGKSPRRRAASMDNNSKFAKSRGR
                     AAKKKASLQSGQEGAGDSPGSQFSKWPASPGSHSNDDFDNWSTFRPRTSSNASTISGR
                     LSPIMTEQDDLGNGDVHSMVYPPSAAKMASTLPSLSEISNPENMENLLDNLNLLSSPT
                     SLTVSTQSSPGTIMQQTPCYSFAPPNTSLNSPSPNYQKYTYGQSSMSPLPQMPMQTLQ
                     DSKSSYGGMAQYNCAAGLLKELLTSDSPPHNDIMTPVDPGVAQPNSRVLGQNVLMGPS
                     SVMPAYGGQASHNKMMNPSHSHPGHGQSTSAVNGRALPHAVNTMPHASGMNRLAQEK
                     TALQVPLPHPMQMNALGGYSPASTCNGYGRMGLLHQEKLPSDLDGMFIERLDCDMESI
                     IRNDLMDGDTLDFNFDNVLPNQSFPHSVKTTHSWVSG"
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:                 33.6          Length:         2162
Percent Similarity:    10.00%        Matches:          10
Best Local Similarity: 100.0%        Conservative:      0
Query Match:           100.0%        Mismatches:        0
DB:                     4.7%         Indels:            0
                        4            Gaps:              0
```

```
US-09-989-890-238  (1-212) x AY094061  (1-2162)

Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||||
Db     198 GCGGGCCTGCCCTCGGCCTCGGCGGCCGCT 227
```

```
RESULT 28
BC021981
LOCUS       BC021981              2413 bp    mRNA    linear   PRI 09-JUL-2005
DEFINITION  Homo sapiens forkhead box O1A (rhabdomyosarcoma), mRNA (cDNA clone
            MGC:1750 IMAGE:2959021), complete cds.
ACCESSION   BC021981
VERSION     BC021981.2  GI:33869892
KEYWORDS    MGC.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2413)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  CONSRTM   Mammalian Gene Collection Program Team
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED    12477932
REFERENCE   2  (bases 1 to 2413)
  CONSRTM   NIH MGC Project
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Bethesda, MD 20892-2590, USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     On Aug 19, 2003 this sequence version replaced gi:18314374.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 2 Row: o Column: 10.
FEATURES             Location/Qualifiers
     source          1..2413
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:1750 IMAGE:2959021"
                     /tissue_type="Lymph, Burkitt lymphoma"
                     /clone_lib="NIH_MGC_8"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1..2413
                     /gene="FOXO1A"
                     /note="synonyms: FOXO1, FKH1"
                     /db_xref="GeneID:2308"
                     /db_xref="MIM:136533"
     CDS             314..2281
                     /gene="FOXO1A"
                     /codon_start=1
                     /product="forkhead box O1A"
                     /protein_id="AAH21981.1"
                     /db_xref="GI:18314375"
                     /db_xref="GeneID:2308"
```

```
              /db_xref="MIM:136533"
              /translation="MAEAPQVVEIDDPDFEPLPRPRSCTWPLPRPEFSQSNSATSSPAP
              SGSAAANPDAAAGLPSASAAAVSADFMSNLSLLEESEDFPQAPGSVAAAVAAAAAAA
              TGGLCGDFQGPEAGCLHPAPPQPPPGPLSQHPPVPAAAGPLAGOPRKSSSSRRNAW
              GNLSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLS
              LHSKFIRVQNEGTGKSSWWMLNPEGGKSGKSPRRAASMDNNSKFAKSRRAAKKKAS
              LQSGQEGAGDSPGSQFSKWPASPGSHSNDDFDNWSTFRPRTSSNASTISGRLSPIMTE
              QDDLGEGDVHSMVYPPSAAKMASTLPSLSEISNPENMENLLDLNLLSSPTSLTVSTQ
              SSPGTMWQQTPCYSFAPPNTSLNSPSPNYQKYTYGQGSMSPLPQMPIQTLQDNKSSYG
              GMSQYNCAPGLLKELLTSDSPPHNDIMTPVDPGVAQPNSRVLGQNVMMGPNSVMSTYG
              SQASHNKMMNPSSHTHPGHAQQTSAVNGRPLPHTVSTMPHTSGMNRLTQVKTPVQVPL
              PHPMQMSALGGYSSVSSCNGYGRMGLLHQEKLPSDLDGMFIERLDCDMESIIRNDLMD
              GDTLDFNFDNVLPNQSFPHSVKTTTHSWVSG"
ORIGIN
Alignment Scores:
Pred. No.:         36.9         Length:          2413
Score:             10.00        Matches:         10
Percent Similarity: 100.0%      Conservative:    0
Best Local Similarity: 100.0%   Mismatches:      0
Query Match:       4.7%         Indels:          0
DB:                8            Gaps:            0

US-09-989-890-238 (1-212) x BC021981 (1-2413)

Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||||
Db     479 GCGGGACCTGCCCCTCGGCCCTGCCGCCGCT 508

RESULT 29
BC070065
LOCUS      BC070065        2482 bp    mRNA    linear   PRI 21-JUL-2005
DEFINITION Homo sapiens forkhead box O1A (rhabdomyosarcoma), mRNA (cDNA clone
           IMAGE:30345006), partial cds.
ACCESSION  BC070065
VERSION    BC070065.1  GI:47123315
KEYWORDS   .
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Hominidae; Homo.
REFERENCE  1  (bases 1 to 2482)
AUTHORS    Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
           Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
           Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
           Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
           Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
           Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
           Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
           Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
           Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
           McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
           Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
           Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
           Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
           Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
           Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
           Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
           Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
           Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE      Generation and initial analysis of more than 15,000 full-length
           human and mouse cDNA sequences
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED     12477932
REFERENCE  2  (bases 1 to 2482)
AUTHORS    NIH MGC Project
CONSRTM    NIH MGC Project
TITLE      Direct Submission
JOURNAL    Submitted (10-MAY-2004) National Institutes of Health, Mammalian
           Gene Collection (MGC), Bethesda, MD 20892-2590, USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
```

```
COMMENT    Contact: MGC help desk
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Dr. Stefan Hansson
           cDNA Library Preparation: Michael Brownstein / Ted Usdin
           Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Sequencing Group at the Stanford Human Genome
           Center, Stanford University School of Medicine, Stanford, CA  94305
           Web site:       http://www.shgc.stanford.edu
           Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
           Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
           R. M.

           Clone distribution: MGC clone distribution information can be found
           through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
           Series: IRAK Plate: 168 Row: n Column: 5
           This clone was selected for full length sequencing because it
           passed the following selection criteria: matched mRNA gi: 9257221.
FEATURES             Location/Qualifiers
     source          1..2482
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30345006"
                     /tissue_type="Placenta, normal"
                     /clone_lib="NIH_MGC_147"
                     /lab_host="DH10B"
                     /note="Vector: pBluescriptR"
     gene            <1..>2482
                     /gene="FOXO1A"
                     /note="synonyms: FOXO1, FKH1"
                     /db_xref="GeneID:2308"
                     /db_xref="MIM:136533"
     CDS             <1..>2344
                     /gene="FOXO1A"
                     /codon_start=2
                     /product="FOXO1A protein"
                     /protein_id="AAH70065.2"
                     /db_xref="GI:71051130"
                     /db_xref="GeneID:2308"
                     /db_xref="MIM:136533"
                     /translation="HSTGSSAAGAPLGRASGPRRPSVLPSAALSAGARRLCPGPAAL
                     AGRPVRAADPEEPRCGWPREVKFWARASTPPRLPPSFRPLAAPASFPQISDRPFAPPP
                     RPPPVLRSPPLGSPAAGGGAGVTMAEAPQVVEIDDPDFEPLPRPRSCTWPLPRPEFSQS
                     NSATSSPAPSGSAAANPDAAAGLPSASAAAVSADFMSNLSLLEESEDFPQAPGSVAAA
                     VAAAAAATGGLCGDFQGPEAGCLHPAPPQPPPGPLSQHPPVPAAAGPLAGOPRK
                     SSSSRRNAWGNLSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGW
                     KNSIRHNLSLHSKFIRVQNEGTGKSSWWMLNPEGGKSGKSPRRAASMDNNSKFAKSR
                     SRAAKKKASLQSGQEGAGDSPGSQFSKWPASPGSHSNDDFDNWSTFRPRTSSNASTIS
                     GRLSPIMTEQDDLGEGDMHSMVYPPSAAKMASTLPSLSEISNPENMENLLDLNLLSS
                     PTSLTVSTQSSPGTMWQQTPCYSFAPPNTSLNSPSPNYQKYTYGQGSMSPLPQMPIQT
                     LQDNKSSYGGMSQYNCAPGLLKELLTSDSPPHNDIMTPVDPGVAQPNSRVLGQNVMMG
                     PNSVMSTYGSQASHNKMMNPSSHTHPGHAQQTSAVNGRPLPHTVSTMPHTSGMNRLTQ
                     VKTPVQVPLPHPMQMSALGGYSSVSSCNGYGRMGLLHQEKLPSDLDGMFIERLDCDME
                     SIIRNDLMDGDTLDFNFDNVLPNQSFPHSVKTTTHSWVSG"
ORIGIN

Alignment Scores:
Pred. No.:         37.8         Length:          2482
Score:             10.00        Matches:         10
Percent Similarity: 100.0%      Conservative:    0
Best Local Similarity: 100.0%   Mismatches:      0
Query Match:       4.7%         Indels:          0
DB:                8            Gaps:            0

US-09-989-890-238 (1-212) x BC070065 (1-2482)

Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||||
Db     542 GCGGGACCTGCCCCTCGGCCCTGCCGCCGCT 571

RESULT 30
CQ843453/c
```

LOCUS   CQ843453   2816 bp   DNA   linear   PAT 02-AUG-2004
DEFINITION   Sequence 2100 from Patent EP1440981.
ACCESSION   CQ843453
VERSION   CQ843453.1 GI:50895240
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE   1
AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
                Yamamoto,J., Isono,Y., Nagai,K. and Irie,R.
TITLE   Full-length human cdna
JOURNAL   Patent: EP 1440981-A 2100 28-JUL-2004;
                Research Association for Biotechnology (JP)
FEATURES   Location/Qualifiers
     source   1..2816
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:   42.1   Length: 2816
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x CQ843453 (1-2816)

Qy   26 SerSerLeuAlaLeuSerArgProLeu 35
             |||||||||||||||||||||||||||
Db   261 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 232

RESULT 31
AK126430/c
LOCUS   AK126430   2816 bp   mRNA   linear   PRI 09-SEP-2003
DEFINITION   Homo sapiens cDNA FLJ44466 fis, clone UTERU2025645.
ACCESSION   AK126430
VERSION   AK126430.1 GI:34532909
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE   1
AUTHORS   Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
                Kodaira,H., Furuya,T., Takahashi,M., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
                Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
                Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
                Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
                Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
                Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
                Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
                Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2816)
AUTHORS   Isogai,T. and Yamamoto,J.
TITLE   Direct Submission
JOURNAL   Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
                Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
                (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
                Economy, Trade and Industry of Japan; cDNA full insert sequencing:
                Research Association for Biotechnology (RAB); cDNA library
                construction: Helix Research Institute (HRI) (supported by Japan
                Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
                HRI, and Biotechnology Center, National Institute of Technology and

Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES   Location/Qualifiers
     source   1..2816
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UTERU2025645"
                     /tissue_type="uterus"
                     /clone_lib="UTERU2"
CDS   263..1021
                     /note="cloning vector: pME18SFL3"
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC86548.1"
                     /db_xref="GI:34532910"
                     /translation="MPTIQALCLLPKVSTPSITVPSPQRSAALSITVPSPQRSAAPSI
                     TVPSPQRSAPRASLCPPHKGQHPEHHCALPPKGQHPEHHCALPTKVSTPSITVPSPQR
                     SPQMSLSSAARGSNTDVAGLSVGEWPGWLWGEGQDGAQQRPHLPSGSGAGVAPQRL
                     PKSRACIICSRHGAQGPWVTGRSVSHSSHCPIQSLLDLQRPDLGTDWGRTRPLCTPQDL
                     CGGRPLPSTWGVTMHIHCLSVSLSLCLSLSHCV"
ORIGIN

Alignment Scores:
Pred. No.:   42.1   Length: 2816
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x AK126430 (1-2816)

Qy   26 SerSerLeuAlaLeuSerArgProLeu 35
             |||||||||||||||||||||||||||
Db   261 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 232

RESULT 32
STMSAM/c
LOCUS   STMSAM   2981 bp   DNA   linear   BCT 02-SEP-1997
DEFINITION   Streptomyces coelicolor DNA for aspartate
                aminotransferase,ribosomal protein, partial and complete cds.
ACCESSION   D32254
VERSION   D32254.1 GI:971285
KEYWORDS   nusG; secE; rplK; rplA; ribosomal protein; aspartate
                aminotransferase.
SOURCE   Streptomyces coelicolor A3(2)
ORGANISM   Streptomyces coelicolor A3(2)
                Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
                Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1 (bases 1 to 2981)
AUTHORS   Puttikhunt,C., Nihira,T. and Yamada,Y.
TITLE   Cloning, nucleotide sequence, and transcriptional analysis of the
                nusG gene of Streptomyces coelicolor A3(2), which encodes a
                putative transcriptional antiterminator
JOURNAL   Mol. Gen. Genet. 247 (1), 118-122 (1995)
PUBMED   7715599
REFERENCE   2 (bases 1 to 2981)
AUTHORS   Puttikhunt,C.
TITLE   Direct Submission
JOURNAL   Submitted (20-JUL-1994) Chunya Puttikhunt, Osaka University,
                Department of Biotechnology; 2-1 Yamadaoka, Suita, Osaka 565, Japan
                (Tel:06-877-5111(ex.3441), Fax:06-879-7448)
COMMENT   Submitted (20-Jul-1994) to DDBJ by:
                Puttikhunt Chunya
                Osaka University
                Department of Biotechnology
                2-1 Yamadaoka
                Suita, Osaka 565
                Japan
                Phone: 06-877-5111 x3441
                Fax: 06-879-7448.
FEATURES   Location/Qualifiers

```
source        1..2981
              /organism="Streptomyces coelicolor A3(2)"
              /mol_type="genomic DNA"
              /isolate="A3(2)"
              /db_xref="taxon:100226"
CDS           complement(<1..368)
              /citation=[1]
              /codon_start=1
              /evidence=not_experimental
              /product="Aspartate aminotransferase"
              /protein_id="BAA21035.1"
              /db_xref="GI:2351224"
              /translation="MSAAIPPTERRVSARVGAISESATLAVDAKAKALKAAGRPVIGF
              GAGEPDFPTPDYIVEAAVEACKNPKFHRYTPAGGLPELKAAIAAKTLRDSGYEVDPSQ
              ILVTNGGKQAIYEAFRAILDP"
RBS           782..785
              /citation=[1]
              /evidence=not_experimental
gene          789..1073
              /gene="secE"
CDS           789..1073
              /gene="secE"
              /function="secretion"
              /citation=[1]
              /codon_start=1
              /evidence=not_experimental
              /product="Secretion"
              /protein_id="BAA06984.1"
              /db_xref="GI:971287"
              /translation="MTDAVGSIDMPDAQDEAPDSKKSRKGGKRGKKGPLKRLALFYRQ
              IVADVRKVVWFSRNQLTTYTTVVIIFVVIMIGLVTLIDYGFSHAAKYVFG"
RBS           1137..1140
              /citation=[1]
              /evidence=not_experimental
gene          1154..2056
              /gene="nusG"
CDS           1154..2056
              /gene="nusG"
              /citation=[1]
              /codon_start=1
              /evidence=not_experimental
              /protein_id="BAA06985.1"
              /db_xref="GI:971288"
              /translation="MSDPNVNDAIEPVESVEDELGTVEGADNEDTEASARAEAAADDTV
              VAETDEDATDAEADETAEATDAEATEABADADADAEAEESSEEPEAREEPELDPIEKL
              RQDVRVLPGEWVIHTYAGYENRVKTNLEQRAVSLNVEDYIFQAEVPQEEVVQIKNGD
              RKTIRQNKLPGVVLVRMDLTNESWGVVRNTPGVTGFVGNAYDPYPLTLDEIVKMLAPE
              AEEKAAREAAEAEGKPAPQRKVEVQVLDFEVGDSVTVTDGPFATLQATINEINADSKK
              VKGLVEIFGRETPVELSFDQIQKN"
RBS           2229..2232
              /citation=[1]
              /evidence=not_experimental
gene          2244..2678
              /gene="rplK"
CDS           2244..2678
              /gene="rplK"
              /citation=[1]
              /codon_start=1
              /evidence=not_experimental
              /product="ribosomal protein"
              /protein_id="BAA06986.1"
              /db_xref="GI:971289"
              /translation="MPPKKKVTGLIKLQIQAGAANPAPPVGPALGQHGVNIMEFCKA
              YNAATESQRGWVIPVEITVYEDRSFTFITKTPAAKMILKAAGVEKGSGEPHKTKVAK
              ITRDQVRBIATTKMPDLNANDLDQAEKIIAGTARSMGVTVEG"
RBS           2752..2755
              /citation=[1]
              /evidence=not_experimental
gene          2762..2980
              /gene="rplA"
```

```
CDS           2762..>2980
              /gene="rplA"
              /citation=[1]
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="ribosomal protein"
              /protein_id="BAA22059.1"
              /db_xref="GI:971290"
              /translation="MSKRSKSLRAADAKIDRDKLYAPLEAVRLAKETTSKFDGTVBV
              AFRLGVDPRKADQMVRGTVNLPHGTGKTA"
ORIGIN

Alignment Scores:
Pred. No.:                         2981
Score:                  44.2       Matches:           10
Percent Similarity:     10.00%     Conservative:      0
Best Local Similarity:  100.0%     Mismatches:        0
Query Match:            100.0%     Indels:            0
DB:                     4.7%       Gaps:              0
                        1

US-09-989-890-238 (1-212) x STMSAM (1-2981)

Qy     51 GlyAlaGlyLeuProSerAlaSerAla 60
          |||||||||||||||||||||||||||
Db   1860 GGCGCCGGCTTGCCCTCGGCCTCGGCGGCC 1831

RESULT 33
CQ874011
LOCUS        CQ874011                3421 bp    DNA      linear   PAT 27-SEP-2004
DEFINITION   Sequence 430 from Patent WO2004076622.
ACCESSION    CQ874011
VERSION      CQ874011.1  GI:5747603
KEYWORDS     .
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE    1
  AUTHORS    Taira,K. and Kawasaki,H.
  TITLE      Regulation of mammalian cells
  JOURNAL    Patent: WO 200407622-A 430 10-SEP-2004;
             National Institute of Advanced Industrial Science and Tec hnology
             (JP)
FEATURES             Location/Qualifiers
     source          1..3421
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                         3421
Score:                  49.8       Matches:           10
Percent Similarity:     10.00%     Conservative:      0
Best Local Similarity:  100.0%     Mismatches:        0
Query Match:            100.0%     Indels:            0
DB:                     4.7%       Gaps:              0
                        6

US-09-989-890-238 (1-212) x CQ874011 (1-3421)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db    172 GCGGGCCTGCCCTCGGCCTCGGCTGCCGCT 201

RESULT 34
HSU02310
LOCUS        HSU02310                3421 bp    mRNA     linear   PRI 16-DEC-1993
DEFINITION   Human fork head domain protein (FKHR) mRNA, complete cds.
ACCESSION    U02310
VERSION      U02310.1  GI:435422
KEYWORDS     .
```

```
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 3421)
AUTHORS     Galli,N.
TITLE       Fusion of a fork head domain to PAX-3 in the solid tumor alveolar
            rhabdomyosarcoma
JOURNAL     Nature Genetics 5(3), 230-235 (1993)
REFERENCE   2 (bases 1 to 3421)
AUTHORS     Galli,N., Davis,R.J., Fredericks,W.J., Mukhopadhyay,S.,
            Rauscher,F.J. III, Emanuel,B.S., Rovera,G. and Barr,F.G.
TITLE       Direct Submission
JOURNAL     Submitted (01-OCT-1993) Naomi Galli, Wistar Institute, 3606 Spruce
            Street, Philadelphia, PA 19104 USA
FEATURES    Location/Qualifiers
    source  1..3421
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
    gene    1..3421
            /gene="FKHR"
    CDS     7..1974
            /gene="FKHR"
            /codon_start=1
            /product="fork head domain protein"
            /protein_id="AAA03629.1"
            /db_xref="GI:433423"
            /translation="MAEAPQVVEIDPDPEPLPRPRSCTWPLPRPEFSQSNSATSSPAP
            SGSAAANPDAAAGLPSASAAAVSADFMSNLSLLEESEDFPQAPGSVAAAVAAAAAAAA
            TGGLCLGDFQGPBAGCLHPAPPQPPPPGPVSQHPPVVPAAAGPLAGQPRKSSSSRRNAW
            GNLSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLS
            LHSKFIRVQNEGTGKSSWWMLNPEGGKSGKSPRRRAASMDNNSKFAKSRSRAAKCKAS
            LQSGQGPAGDSPGSQSFKWPASPGSHSNDDFDNWSTFRPRTSSNASTISGRLSPIMTE
            QDDLGBGDVASMVYPPSAAKMASTLPSLSEISNPENMENLLDNLNLLSSPTSLTVSTQ
            SSPGTMMQQTPCYSFAPPNTSLNSPSNYQKYTYGQSSMSPLPQMPIQTLQDNKSSYG
            GMSQYNCAPGLLKELITSDSPHNDIMTPVDFPGVAQPNSRVLGQNVMMGPNSVMSTYG
            SQASHNKMNPSHHTHPGHAQQTSAVNGRPLPHTVSTMPHTSGMNRLTQVKTPVQVPL
            PHPMQMSALGGYSSVSCNGYGRMGLLHQEKLPSDLDGMFIERLLDCDMESIIRNDLMD
            GDTLDFNFDNVLPNQSFPHSVKTTTHSWVSG"
ORIGIN

Alignment Scores:
Pred. No.:              49.8              Length:            3421
Score:                  10.00            Matches:           10
Percent Similarity:     100.0%           Conservative:      0
Best Local Similarity:  100.0%           Mismatches:        0
Query Match:            4.7%             Indels:            0
DB:                     8                Gaps:              0

US-09-989-890-238 (1-212) x HSU02310 (1-3421)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db    172 GCGGGCCTGCCCTCGGCTGCCTCCGCGCT 201

RESULT 35
CS031181
LOCUS       CS031181            5723 bp    DNA       linear   PAT 10-MAR-2005
DEFINITION  Sequence 687 from Patent WO2005016962.
ACCESSION   CS031181
VERSION     CS031181.1  GI:60731351
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
AUTHORS     Abbas,A., Clark,H., Ouyang,W., Williams,M.P., Wood,W.I. and Wu,T.D.
TITLE       Compositions and methods for the treatment of immune related
            diseases
JOURNAL     Patent: WO 2005016962-A 687 24-FEB-2005;
            Genentech, Inc. (US)
FEATURES    Location/Qualifiers
    source  1..5723
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                               Length:            5723
Score:                  77.4             Matches:           10
Percent Similarity:     10.00            Conservative:      0
Best Local Similarity:  100.0%           Mismatches:        0
Query Match:            100.0%           Indels:            0
DB:                     4.7%             Gaps:              0
                        6

US-09-989-890-238 (1-212) x CS031181 (1-5723)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db    551 GCGGGCCTGCCCTCGGCTGCCTCCGCGCT 580

RESULT 36
CS040133
LOCUS       CS040133            5723 bp    DNA       linear   PAT 22-MAR-2005
DEFINITION  Sequence 687 from Patent WO2005019258.
ACCESSION   CS040133
VERSION     CS040133.1  GI:61847976
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
AUTHORS     Abbas,A., Clark,H., Ouyang,W., Williams,P.M., Wood,W.I. and Wu,T.D.
TITLE       Compositions and methods for the treatment of immune related
            diseases
JOURNAL     Patent: WO 2005019258-A 687 03-MAR-2005;
            Genentech, Inc. (US)
FEATURES    Location/Qualifiers
    source  1..5723
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                               Length:            5723
Score:                  77.4             Matches:           10
Percent Similarity:     10.00            Conservative:      0
Best Local Similarity:  100.0%           Mismatches:        0
Query Match:            100.0%           Indels:            0
DB:                     4.7%             Gaps:              0
                        6

US-09-989-890-238 (1-212) x CS040133 (1-5723)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db    551 GCGGGCCTGCCCTCGGCTGCCTCCGCGCT 580

RESULT 37
AX587664
LOCUS       AX587664            5723 bp    DNA       linear   PAT 10-JAN-2003
DEFINITION  Sequence 134 from Patent WO0246467.
ACCESSION   AX587664
VERSION     AX587664.1  GI:28212364
KEYWORDS    .
SOURCE      synthetic construct
ORGANISM    synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
```

```
AUTHORS    Bertucci,F., Houlgatte,R., Birnbaum,D., Nguyen,C., Viens,P. and
           Pert,V.
TITLE      Gene expression profiling of primary breast carcinomas using arrays
           of candidate genes
JOURNAL    Patent: WO 0246467-A 134 13-JUN-2002;
           Ipsogen (FR)
FEATURES             Location/Qualifiers
     source          1..5723
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="primer"
     misc_feature    1..5723
                     /note="forkhead box o1a (rhabdomyosarcoma) (FOXO1A) gene."
ORIGIN

Alignment Scores:
Pred. No.:                      77.4           Length:          5723
Score:                          10.00          Matches:         10
Percent Similarity:             100.0%         Conservative:    0
Best Local Similarity:          100.0%         Mismatches:      0
Query Match:                    4.7%           Indels:          0
DB:                             6              Gaps:            0

US-09-989-890-238 (1-212) x AX587664 (1-5723)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAlaAla 61
         |||||||||||||||||||||||||||||||||
Db   551 GCGGGCCTGCCCTCCGGCCTCGGCTGCCGCT 580

RESULT 38
AF032885
LOCUS       AF032885                5723 bp    mRNA         linear   PRI 19-FEB-1998
DEFINITION  Homo sapiens forkhead protein (FKHR) mRNA, complete cds.
ACCESSION   AF032885
VERSION     AF032885.1  GI:2895491
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 5723)
  AUTHORS   Anderson,M.J., Viars,C.S., Czekay,S., Cavenee,W.K. and Arden,K.C.
  TITLE     Cloning and characterization of three human forkhead genes that
            comprise an FKHR-like gene subfamily
  JOURNAL   Genomics 47 (2), 187-199 (1998)
   PUBMED   9479491
REFERENCE   2  (bases 1 to 5723)
  AUTHORS   Anderson,M.J., Viars,C.S., Czekay,S., Cavenee,W.K. and Arden,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1997) Medicine, Ludwig Institute for Cancer
            Research, San Diego Branch, 9500 Gilman Drive, La Jolla, CA
            92093-0660, USA
FEATURES             Location/Qualifiers
     source          1..5723
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="13q13-q14.1"
     gene            1..5723
                     /gene="FKHR"
     CDS             386..2353
                     /gene="FKHR"
                     /codon_start=1
                     /product="forkhead protein"
                     /protein_id="AAC39591.1"
                     /db_xref="GI:2895492"
                     /translation="MAEAPQVVEIDPDFEPLRPRSCTWPLPRPEFSQSNSATSSPAP
                     SGSAAANPDAAAGLPSASAAVSADFMSNLSLLEESEDFPQAPGSVAAVAAAAAAAA
                     TGGLCGDFQGPEAGCLHAPPQPPPGPLSQHPPVPPAAAGPLAGQPRKSSSSRRNAW
                     GNLSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLS
```

```
                     LHSKFIRVQNEGTGKSSWWMLNPEGGKSGKSPRRAASMDNNSKFAKSRSRAAKKKAS
                     LQSGSGEGAGDSPGSQFSKWPASPGSHSNDDFDNWSTFRPRTSSNASTISGRLSPIMTE
                     QDDLGEGDVHSMVYPPSAAKMASTLPSLSBISNPBNMENLLDNLNLLSSPTSLITVSTQ
                     SSPGTMWQGTPCYSFAPPNTSLNSPSPNYQKYTYGQSSMSPLPQMPIQTLQDNKSSYG
                     GMSQYNCAPGLLKELLTSDPHNDIMTPVDPQVAQPNSRVLGQNVMMGPNSVMSTYG
                     SQASHNKMNPSSHTHPGHAQQTSAVNGRPLPHTVSTMPHTSGMNRLTQVKTPVQVPL
                     PHPMQMSALGGYSVSSCNGYGRMGLLHQEKLPSDLDGMPIERLDCDMBSIIRNDLMD
                     GDTLDFNFDNVLPNQSFPHSVKTTTHSWVSG"
ORIGIN

Alignment Scores:
Pred. No.:                      77.4           Length:          5723
Score:                          10.00          Matches:         10
Percent Similarity:             100.0%         Conservative:    0
Best Local Similarity:          100.0%         Mismatches:      0
Query Match:                    4.7%           Indels:          0
DB:                             8              Gaps:            0

US-09-989-890-238 (1-212) x AF032885 (1-5723)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         |||||||||||||||||||||||||||
Db   551 GCGGGCCTGCCCTCCGGCCTCGGCTGCCGCT 580

RESULT 39
AX779990
LOCUS       AX779990                5769 bp    DNA          linear   PAT 14-JUL-2003
DEFINITION  Sequence 2147 from Patent WO03039443.
ACCESSION   AX779990
VERSION     AX779990.1  GI:33696984
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
            Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.
  TITLE     Novel genetic markers for leukemias
  JOURNAL   Patent: WO 03039443-A 2147 15-MAY-2003;
            Deutsches Krebsforschungszentrum (DE);
            Ludwig-Maximilian-Universitaet Muenchen (DE) ; Haferlach, Torsten,
            PD Dr. Dr. (DE) ; Schoch, Claudia (DB) ; Kern, Wolfgang (DE)
FEATURES             Location/Qualifiers
     source          1..5769
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                      77.9           Length:          5769
Score:                          10.00          Matches:         10
Percent Similarity:             100.0%         Conservative:    0
Best Local Similarity:          100.0%         Mismatches:      0
Query Match:                    4.7%           Indels:          0
DB:                             6              Gaps:            0

US-09-989-890-238 (1-212) x AX779990 (1-5769)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         |||||||||||||||||||||||||||
Db   551 GCGGGCCTGCCCTCCGGCCTCGGCTGCCGCT 580

RESULT 40
SGNUSG/c
LOCUS       SGNUSG                  7235 bp    DNA          linear   BCT 18-APR-2005
DEFINITION  S.griseus nusG, rplKAJL gene cluster.
ACCESSION   X72787
VERSION     X72787.1  GI:575399
KEYWORDS    aspartate aminotransferase; nusG gene; NusG protein;
            rho-independent transcription termination; ribosomal protein L1;
```

```
            ribosomal protein L10; ribosomal protein
            L12; rplA gene; rplJ gene; rplK gene; rplL gene; secE gene;
            tRNA-Trp gene.
SOURCE      Streptomyces griseus
  ORGANISM  Streptomyces griseus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 5960)
  AUTHORS   Kuberski,S., Kasberg,T. and Distler,J.
  TITLE     The nusG gene of Streptomyces griseus: cloning of the gene and
            analysis of the A-factor binding properties of the gene product
  JOURNAL   FEMS Microbiol. Lett. 119 (1-2), 33-39 (1994)
  PUBMED    8039667
REFERENCE   2
  AUTHORS   Kuster,C., Piepersberg,W. and Distler,J.
  TITLE     Cloning and transcriptional analysis of the rplKA-or f31-rplJL gene
            cluster of Streptomyces griseus
  JOURNAL   Mol. Gen. Genet. 257 (2), 219-229 (1998)
  PUBMED    9491081
REFERENCE   3  (bases 1 to 5960)
  AUTHORS   Distler,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1993) J. Distler, Bergische Universitaet,
            Gesamthochschule Wuppertal, Gauss Str 20, 5600 Wuppertal 1, FRG
COMMENT     On Nov 28, 1994 this sequence version replaced gi:288459.
FEATURES             Location/Qualifiers
     source          1..7235
                     /organism="Streptomyces griseus"
                     /mol_type="genomic DNA"
                     /strain="N2-3-11"
                     /db_xref="taxon:1911"
     gene            complement(<1..640)
                     /gene="aatA"
     CDS             complement(<1..640)
                     /gene="aatA"
                     /codon_start=1
                     /transl_table=11
                     /product="aspartate aminotransferase"
                     /protein_id="CAA51294.1"
                     /db_xref="GI:575400"
                     /db_xref="GOA:P36692"
                     /db_xref="InterPro:IPR001176"
                     /db_xref="InterPro:IPR004838"
                     /db_xref="InterPro:IPR004839"
                     /db_xref="UniProt/Swiss-Prot:P36692"
                     /translation="MSAATSPSERRVSARIGAISESATLAVDAKAKALKAAGRPVIGF
                     GAGEPDFPTPDYIVDAAVEACRNPKYHRYTPQRAPELKAAIAEKTLRDSGYEVDAGQI
                     LVTNGGKQAIYEAFAAILDPGDEVIVPAPYWTYPESIRLAGGVPVEVADETTGYRV
                     SVEQLEAARTEKTKVVLFVSPSNPTGAVYSEADAEAIGRWAVEHGLWVMTDEI"
     gene            776..2274
                     /gene="secE/tRNA/nusG"
     mRNA            776..2274
                     /gene="secE/tRNA/nusG"
     gene            853..924
                     /gene="tRNA-Trp"
     tRNA            853..924
                     /gene="tRNA-Trp"
                     /product="tRNA-Trp"
     gene            1021..1023
                     /gene="secE"
     RBS             1021..1023
                     /gene="secE"
     gene            1032..1319
                     /gene="secE"
     CDS             1032..1319
                     /gene="secE"
                     /codon_start=1
                     /transl_table=11
                     /product="secretory protein SecE"
                     /protein_id="CAA51295.1"
                     /db_xref="GI:603588"
                     /db_xref="GOA:P36690"
                     /db_xref="InterPro:IPR001901"
                     /db_xref="InterPro:IPR005807"
                     /db_xref="InterPro:IPR005808"
                     /db_xref="UniProt/Swiss-Prot:P36690"
                     /translation="MTDAVGSIDMPDAEDEAPESKKKSRKGGKRGKKGPLGRLALFYR
                     QIVAELRKVVWPTRSQLTTYTSVIVFVVMIGLVTVLDIGFARVVKVFG"
     gene            1381..1384
                     /gene="nusG"
     RBS             1381..1384
                     /gene="nusG"
     gene            1393..2277
                     /gene="nusG"
     CDS             1393..2277
                     /gene="nusG"
                     /codon_start=1
                     /transl_table=11
                     /product="NusG"
                     /protein_id="CAA51296.1"
                     /db_xref="GI:603589"
                     /db_xref="GOA:P36260"
                     /db_xref="InterPro:IPR001062"
                     /db_xref="InterPro:IPR005824"
                     /db_xref="InterPro:IPR006645"
                     /db_xref="InterPro:IPR006646"
                     /db_xref="InterPro:IPR008991"
                     /db_xref="UniProt/Swiss-Prot:P36260"
                     /translation="MSDPNLNDAVEPEAGASESAKDELDIVEAADSVDPDQAEAADLA
                     AGEPAERAAVNVAGDDSDEDDAAABEAVEADDESADEBEABPAAPVDPVAALRDELRT
                     LPGEWVVIHTYAGYEKRVKANLEQRAVSLNVEBFIYQAEVPEEIVQIKNGERKNVRQ
                     NKLPGYVLVRMDLTNESWGVVRNTPGVTGFVGNAYDPYPLTLDEIVKMLAPEAEEKAA
                     REAABAEGKPAPARKVEVQVLDFEVGDSVTVDGPFATLQATINEINADSKKVKGLVE
                     IFGRETPVELSFDQIQKN"
     misc_structure  2309..2357
                     /note="rho-factor independent termination strucure"
     gene            2389..3688
                     /gene="rplK"
     mRNA            2389..3688
                     /gene="rplK"
     RBS             2444..2447
                     /gene="rplK"
     gene            2454..2888
                     /gene="rplL"
     CDS             2454..2888
                     /gene="rplL"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosomal protein L11"
                     /protein_id="CAA51297.1"
                     /db_xref="GI:288461"
                     /db_xref="GOA:P36258"
                     /db_xref="InterPro:IPR000911"
                     /db_xref="InterPro:IPR006519"
                     /db_xref="UniProt/Swiss-Prot:P36258"
                     /translation="MPPKKKVTGLIRKLQINAGAANPAPPVGPALGQHGVNIMEFCKA
                     YNAATESQRGMVVPVEITVYEDRSFTFVTKTPPAAKLILKAAGVDKGSGEPHKTKVAK
                     LTAAQVREIATKLPDLNANDLDAASKIIAGTARSMGITVEG"
     gene            2959..2962
                     /gene="rplA"
     RBS             2959..2962
                     /gene="rplA"
     gene            2969..3691
                     /gene="rplA"
     CDS             2969..3691
                     /gene="rplA"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosomal protein L1"
                     /protein_id="CAA51298.1"
                     /db_xref="GI:1333827"
                     /db_xref="GOA:P36256"
                     /db_xref="InterPro:IPR002143"
                     /db_xref="InterPro:IPR005878"
                     /db_xref="UniProt/Swiss-Prot:P36256"
                     /translation="MKRSKNLRAADAKVDRERNYAPLEAVRLAKETSSTKFDGTVEVA
```

```
gene              FALGVDPRKADQMVRGTVNLPHGTGKTARVLVFATGDRAAAEAAGADIVGADELIDE
                  VAKGRLDFDAVVATPDLMGKVGRLGRVLGPRGLMPNPKTGTVTPDVVKAVNDIKGGKI
                  EFRVDKHSNLHFIIGKVSFDETKLVENYAAALEEILRLKPSAAKGRYIKKATLATTMG
                  PGIPLDANRTRNLLVEEDPASV"
                  3786. .4779
      /gene="orf31"
mRNA              3786. .4776
      /gene="orf31"
                  /note="putative lipoprotein"
RBS               3852. .3854
      /gene="orf31"
CDS               3862. .4779
      /gene="orf31"
                  /note="putative lipoprotein"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="CAA51299.1"
                  /db_xref="GI:575402"
                  /db_xref="InterPro:IPR000437"
                  /db_xref="UniProt/Swiss-Prot:P36261"
                  /translation="MRITSARRTGTALAVAAALTSIAACSGSDGAKGSDGAGEGKAGA
                  VSKASPVAALKQVQQKTGGAQSAKVRGTTEMGSVMSMKQSGAIGWADGLSGALTITYT
                  GGTMGDALKQAGDGSVQARYFKDEYYANMGDAMAANTGGKHWIRYSYKDLAELGGAS
                  GDVMKDQIGNSTPDAGGVKALLASGDDVQNSTPATHYSGTVDVAGLTAKNSNL
                  DAEQLAAFKEQLALAGVTTQTVDIWVDKNDLLVKKTERGEMKTGSFNSTIFYSDYGTE
                  VPTEKPAASDTVDFKEMLKQGGATPGATS"
gene              4847. .6077
      /gene="rplJ/rplL"
mRNA              4847. .6077
      /gene="rplJ/rplL"
gene              5039. .5043
      /gene="rplJ"
RBS               5039. .5043
      /gene="rplJ"
gene              5055. .5612
      /gene="rplJ"
CDS               5055. .5612
      /gene="rplJ"
                  /codon_start=1
                  /transl_table=11
                  /product="ribosomal protein L10"
                  /protein_id="CAA51300.1"
                  /db_xref="GI:288464"
                  /db_xref="GOA:P36257"
                  /db_xref="InterPro:IPR001790"
                  /db_xref="InterPro:IPR002363"
                  /db_xref="UniProt/Swiss-Prot:P36257"
                  /translation="MARPDKAAVAELTDQFRSSNAAVLTEYRGLTVAQLKELRRSLG
```

```
Alignment Scores:
Pred. No.:             94.6                 Length:          7235
Score:                 10.00                Matches:         10
Percent Similarity:    100.0%               Conservative:    0
Best Local Similarity: 100.0%               Mismatches:      0
Query Match:           4.7%                 Indels:          0
DB:                    1                    Gaps:            0

US-09-989-890-238 (1-212) x SGNUSG (1-7235)

Qy     51  GlyAlaGlyLeuProSerAlaSerAlaAla  60
           ||||||||||||||||||||||||||||||
Db   2081  GGAGCCGGCTTGCCCTCGGCCTCGGCGGCC  2052

RESULT 41
AC137577/c
LOCUS        AC137577      68170 bp    DNA    linear    HTG  24-NOV-2002
DEFINITION   Homo sapiens chromosome 8 clone RP11-68J16 map 8, LOW-PASS SEQUENCE
             SAMPLING.
ACCESSION    AC137577
VERSION      AC137577.1  GI:25229244
KEYWORDS     HTG; HTGS_PHASE0.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
```

```
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE    1  (bases 1 to 68170)
AUTHORS      Birren,B., Nusbaum,C. and Lander,E.
TITLE        Homo sapiens chromosome 8, clone RP11-68J16
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 68170)
AUTHORS      Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
             Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
             Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
             Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
             Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
             Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
             Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
             Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
             Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
             Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
             Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
             Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
             Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
             Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
             Seaman,S., Severy,P., Smith,C., Spencer,S., Stange-Thomann,N.,
             Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
             Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
             Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE        Direct Submission
JOURNAL      Submitted (24-NOV-2002) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      All repeats were identified using RepeatMasker.
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             ---------  Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
             -----------------  Project Information
               Center project name: L28701
               Center clone name: 68_J_16
             -----------------
             * NOTE: This record contains 86 individual
             * sequencing reads that have not been assembled into
             * contigs. Runs of N are used to separate the reads
             * and the order in which they appear is completely
             * arbitrary. Low-pass sequence sampling is useful for
             * identifying clones that may be gene-rich and allows
             * overlap relationships among clones to be deduced.
             * However, it should not be assumed that this clone
             * will be sequenced to completion. In the event that
             * the record is updated, the accession number will
             * be preserved.
             *                1 683:  contig of 683 bp in length
             *              684 783:  gap of 100 bp
             *              784 1471: contig of 688 bp in length
             *             1472 1571: gap of 100 bp
             *             1572 2247: contig of 676 bp in length
             *             2248 2347: gap of 100 bp
             *             2348 3040: contig of 693 bp in length
             *             3041 3140: gap of 100 bp
             *             3141 3837: contig of 697 bp in length
             *             3838 3937: gap of 100 bp
             *             3938 4625: contig of 688 bp in length
             *             4626 4725: gap of 100 bp
             *             4726 5398: contig of 673 bp in length
             *             5399 5498: gap of 100 bp
             *             5499 6197: contig of 699 bp in length
             *             6198 6297: gap of 100 bp
             *             6298 6987: contig of 690 bp in length
             *             6988 7087: gap of 100 bp
             *             7088 7771: contig of 684 bp in length
             *             7772 7871: gap of 100 bp
             *             7872 8566: contig of 695 bp in length
```

```
*    8567    8666:   gap of 100 bp
*    8667    9352:   contig of 686 bp in length
*    9353    9452:   gap of 100 bp
*    9453   10131:   contig of 679 bp in length
*   10132   10231:   gap of 100 bp
*   10232   10928:   contig of 697 bp in length
*   10929   11028:   gap of 100 bp
*   11029   11718:   contig of 690 bp in length
*   11719   11818:   gap of 100 bp
*   11819   12506:   contig of 688 bp in length
*   12507   12606:   gap of 100 bp
*   12607   13291:   contig of 685 bp in length
*   13292   13391:   gap of 100 bp
*   13392   14084:   contig of 693 bp in length
*   14085   14184:   gap of 100 bp
*   14185   14876:   contig of 692 bp in length
*   14877   14976:   gap of 100 bp
*   14977   15667:   contig of 691 bp in length
*   15668   15767:   gap of 100 bp
*   15768   16467:   contig of 700 bp in length
*   16468   16567:   gap of 100 bp
*   16568   17255:   contig of 688 bp in length
*   17256   17355:   gap of 100 bp
*   17356   18046:   contig of 691 bp in length
*   18047   18146:   gap of 100 bp
*   18147   18850:   contig of 704 bp in length
*   18851   18950:   gap of 100 bp
*   18951   19663:   contig of 713 bp in length
*   19664   19763:   gap of 100 bp
*   19764   20456:   contig of 693 bp in length
*   20457   20556:   gap of 100 bp
*   20557   21261:   contig of 705 bp in length
*   21262   21361:   gap of 100 bp
*   21362   22060:   contig of 699 bp in length
*   22061   22160:   gap of 100 bp
*   22161   22858:   contig of 698 bp in length
*   22859   22958:   gap of 100 bp
*   22959   23651:   contig of 693 bp in length
*   23652   23751:   gap of 100 bp
*   23752   24454:   contig of 703 bp in length
*   24455   24554:   gap of 100 bp
*   24555   25250:   contig of 696 bp in length
*   25251   25350:   gap of 100 bp
*   25351   26043:   contig of 693 bp in length
*   26044   26143:   gap of 100 bp
*   26144   26839:   contig of 696 bp in length
*   26840   26939:   gap of 100 bp
*   26940   27627:   contig of 688 bp in length
*   27628   27727:   gap of 100 bp
*   27728   28440:   contig of 713 bp in length
*   28441   28540:   gap of 100 bp
*   28541   29252:   contig of 712 bp in length
*   29253   29352:   gap of 100 bp
*   29353   30057:   contig of 705 bp in length
*   30058   30157:   gap of 100 bp
*   30158   30852:   contig of 695 bp in length
*   30853   30952:   gap of 100 bp
*   30953   31645:   contig of 693 bp in length
*   31646   31745:   gap of 100 bp
*   31746   32434:   contig of 689 bp in length
*   32435   32534:   gap of 100 bp
*   32535   33217:   contig of 683 bp in length
*   33218   33317:   gap of 100 bp
*   33318   34010:   contig of 693 bp in length
*   34011   34110:   gap of 100 bp
*   34111   34806:   contig of 696 bp in length
*   34807   34906:   gap of 100 bp
*   34907   35598:   contig of 692 bp in length
*   35599   35698:   gap of 100 bp
*   35699   36389:   contig of 691 bp in length
*   36390   36489:   gap of 100 bp
*   36490   37195:   contig of 706 bp in length
*   37196   37295:   gap of 100 bp

*   37296   38001:   contig of 706 bp in length
*   38002   38101:   gap of 100 bp
*   38102   38792:   contig of 691 bp in length
*   38793   38892:   gap of 100 bp
*   38893   39581:   contig of 689 bp in length
*   39582   39681:   gap of 100 bp
*   39682   40376:   contig of 695 bp in length
*   40377   40476:   gap of 100 bp
*   40477   41155:   contig of 679 bp in length
*   41156   41255:   gap of 100 bp
*   41256   41949:   contig of 694 bp in length
*   41950   42049:   gap of 100 bp
*   42050   42749:   contig of 700 bp in length
*   42750   42849:   gap of 100 bp
*   42850   43523:   contig of 674 bp in length
*   43524   43623:   gap of 100 bp
*   43624   44324:   contig of 701 bp in length
*   44325   44424:   gap of 100 bp
*   44425   45124:   contig of 700 bp in length
*   45125   45224:   gap of 100 bp
*   45225   45904:   contig of 680 bp in length
*   45905   46004:   gap of 100 bp
*   46005   46704:   contig of 700 bp in length
*   46705   46804:   gap of 100 bp
*   46805   47510:   contig of 706 bp in length
*   47511   47610:   gap of 100 bp
*   47611   48305:   contig of 695 bp in length
*   48306   48405:   gap of 100 bp
*   48406   49071:   contig of 666 bp in length
*   49072   49171:   gap of 100 bp
*   49172   49861:   contig of 690 bp in length
*   49862   49961:   gap of 100 bp
*   49962   50632:   contig of 671 bp in length
*   50633   50732:   gap of 100 bp
*   50733   51429:   contig of 697 bp in length
*   51430   51529:   gap of 100 bp
*   51530   52233:   contig of 704 bp in length
*   52234   52333:   gap of 100 bp
*   52334   53021:   contig of 688 bp in length
*   53022   53121:   gap of 100 bp
*   53122   53832:   contig of 711 bp in length
*   53833   53932:   gap of 100 bp
*   53933   54630:   contig of 698 bp in length
*   54631   54730:   gap of 100 bp
*   54731   55418:   contig of 688 bp in length
*   55419   55518:   gap of 100 bp
```

```
Alignment Scores:
Pred. No.:                   648              Length:        68170
Score:                    10.00              Matches:          10
Percent Similarity:      100.0%              Conservative:      0
Best Local Similarity:   100.0%              Mismatches:        0
Query Match:               4.7%              Indels:            0
DB:                          14              Gaps:              0

US-09-989-890-238 (1-212) x AC137577 (1-68170)

Qy     25 ProSerLeuAlaLeuLeuSerArgPro 34
                ||||||||||||||||||||||||||
Db  31848 CCGAGCTCCCTGGCTCTACTATCCAGGCCG 31819

RESULT 42
BA000002_16
WPCOMMENT
Sequence split into 17 fragments  LOCUS  BA000002  Accession BA000002   Length   68170
Fragment Name   Begin      End
BA000002_00          1   110000
BA000002_01     100001   210000
BA000002_02     200001   310000
BA000002_03     300001   410000
BA000002_04     400001   510000
BA000002_05     500001   610000
BA000002_06     600001   710000
```

```
BA000002_07     700001     810000
BA000002_08     800001     910000
BA000002_09     900001    1010000
BA000002_10    1000001    1110000
BA000002_11    1100001    1210000
BA000002_12    1200001    1310000
BA000002_13    1300001    1410000
BA000002_14    1400001    1510000
BA000002_15    1500001    1610000
BA000002_16    1600001    1669695
```

Continuation (17 of 17) of BA000002 from base 1600001 (BA000002 Aeropyrum pernix K1 DNA,

```
Alignment Scores:
Pred. No.:               661            Length:          69695
Score:                 10.00            Matches:            10
Percent Similarity:    100.0%           Conservative:        0
Best Local Similarity: 100.0%           Mismatches:          0
Query Match:             4.7%           Indels:              0
DB:                        1            Gaps:                0
```

US-09-989-890-238 (1-212) x BA000002_16 (1-69695)

```
Qy     26  SerSerLeuAlaLeuLeuSerArgProLeu  35
           ||||||||||||||||||||||||||||||
Db  54002  TCGAGCCTGGCGCTCCTCAGTAGACCGTTG  54031
```

RESULT 43
BA000030_59/c
WPCOMMENT
Sequence split into 91 fragments LOCUS BA000030 Accession BA000030

```
Fragment Name   Begin       End
BA000030_00           1    110000
BA000030_01      100001    210000
BA000030_02      200001    310000
BA000030_03      300001    410000
BA000030_04      400001    510000
BA000030_05      500001    610000
BA000030_06      600001    710000
BA000030_07      700001    810000
BA000030_08      800001    910000
BA000030_09      900001   1010000
BA000030_10     1000001   1110000
BA000030_11     1100001   1210000
BA000030_12     1200001   1310000
BA000030_13     1300001   1410000
BA000030_14     1400001   1510000
BA000030_15     1500001   1610000
BA000030_16     1600001   1710000
BA000030_17     1700001   1810000
BA000030_18     1800001   1910000
BA000030_19     1900001   2010000
BA000030_20     2000001   2110000
BA000030_21     2100001   2210000
BA000030_22     2200001   2310000
BA000030_23     2300001   2410000
BA000030_24     2400001   2510000
BA000030_25     2500001   2610000
BA000030_26     2600001   2710000
BA000030_27     2700001   2810000
BA000030_28     2800001   2910000
BA000030_29     2900001   3010000
BA000030_30     3000001   3110000
BA000030_31     3100001   3210000
BA000030_32     3200001   3310000
BA000030_33     3300001   3410000
BA000030_34     3400001   3510000
BA000030_35     3500001   3610000
BA000030_36     3600001   3710000
BA000030_37     3700001   3810000
BA000030_38     3800001   3910000
BA000030_39     3900001   4010000
BA000030_40     4000001   4110000
BA000030_41     4100001   4210000
BA000030_42     4200001   4310000
BA000030_43     4300001   4410000
BA000030_44     4400001   4510000
BA000030_45     4500001   4610000
BA000030_46     4600001   4710000
BA000030_47     4700001   4810000
BA000030_48     4800001   4910000
BA000030_49     4900001   5010000
BA000030_50     5000001   5110000
BA000030_51     5100001   5210000
BA000030_52     5200001   5310000
BA000030_53     5300001   5410000
BA000030_54     5400001   5510000
BA000030_55     5500001   5610000
BA000030_56     5600001   5710000
BA000030_57     5700001   5810000
BA000030_58     5800001   5910000
BA000030_59     5900001   6010000
BA000030_60     6000001   6110000
BA000030_61     6100001   6210000
BA000030_62     6200001   6310000
BA000030_63     6300001   6410000
BA000030_64     6400001   6510000
BA000030_65     6500001   6610000
BA000030_66     6600001   6710000
BA000030_67     6700001   6810000
BA000030_68     6800001   6910000
BA000030_69     6900001   7010000
BA000030_70     7000001   7110000
BA000030_71     7100001   7210000
BA000030_72     7200001   7310000
BA000030_73     7300001   7410000
BA000030_74     7400001   7510000
BA000030_75     7500001   7610000
BA000030_76     7600001   7710000
BA000030_77     7700001   7810000
BA000030_78     7800001   7910000
BA000030_79     7900001   8010000
BA000030_80     8000001   8110000
BA000030_81     8100001   8210000
BA000030_82     8200001   8310000
BA000030_83     8300001   8410000
BA000030_84     8400001   8510000
BA000030_85     8500001   8610000
BA000030_86     8600001   8710000
BA000030_87     8700001   8810000
BA000030_88     8800001   8910000
BA000030_89     8900001   9010000
BA000030_90     9000001   9025608
```

Continuation (60 of 91) of BA000030 from base 5900001 (BA000030 Streptomyces avermitilis

```
Alignment Scores:
Pred. No.:               977            Length:         110000
Score:                 10.00            Matches:            10
Percent Similarity:    100.0%           Conservative:        0
Best Local Similarity: 100.0%           Mismatches:          0
Query Match:             4.7%           Indels:              0
DB:                        1            Gaps:                0
```

US-09-989-890-238 (1-212) x BA000030_59 (1-110000)

```
Qy     51  GlyAlaGlyLeuProSerAlaSerAlaAla  60
           ||||||||||||||||||||||||||||||
Db  65744  GGAGCGGGCTTGCCCTCGGCCTCGGCGGCC  65715
```

RESULT 44
CR555291_0/c
WPCOMMENT
Sequence split into 5 fragments LOCUS CR555291 Accession CR555291

```
Fragment Name   Begin       End
CR555291_0            1    110000
CR555291_1       100001    210000
CR555291_2       200001    310000
```

```
                  CR555291_3      300001    410000
                  CR555291_4      400001    410877
LOCUS       CR555291    410877 bp    DNA    linear   HTG 11-OCT-2004
DEFINITION  Danio rerio clone DKEY-58J8, *** SEQUENCING IN PROGRESS ***, 37
            unordered pieces.
ACCESSION   CR555291
VERSION     CR555291.4  GI:52313314
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 410877)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 17, 2004 this sequence version replaced gi:50724921.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK58J8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 394129 bases at least Q40
            Consensus quality: 399430 bases at least Q30
            Consensus quality: 402692 bases at least Q20
            Insert size: 407277; sum-of-contigs
            Insert size: 200961; 2.9% error; agarose-fp
            Quality coverage: 4.75x in Q20 bases; sum-of-contigs Quality
            coverage: 9.83x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1      5913: contig of 5913 bp in length
            *   5914      6013: gap of 100 bp
            *   6014      8737: contig of 2724 bp in length
            *   8738      8837: gap of 100 bp
            *   8838     12348: contig of 3511 bp in length
            *  12349     12448: gap of 100 bp
            *  12449     34866: contig of 22418 bp in length
            *  34867     34966: gap of 100 bp
            *  34967     50012: contig of 15046 bp in length
            *  50013     50112: gap of 100 bp
            *  50113     52494: contig of 2382 bp in length
            *  52495     52594: gap of 100 bp
            *  52595     55884: contig of 3290 bp in length
            *  55885     55984: gap of 100 bp
            *  55985     59774: contig of 3790 bp in length
            *  59775     59874: gap of 100 bp
            *  59875     71704: contig of 11830 bp in length
            *  71705     71804: gap of 100 bp
            *  71805     90091: contig of 18287 bp in length
            *  90092     90191: gap of 100 bp
            *  90192     94302: contig of 4111 bp in length
            *  94303     94402: gap of 100 bp
            *  94403    104620: contig of 10218 bp in length
            * 104621    104720: gap of 100 bp
            * 104721    110745: contig of 6025 bp in length
            * 110746    110845: gap of 100 bp
            * 110846    113873: contig of 3028 bp in length
            * 113874    113973: gap of 100 bp
            * 113974    118974: contig of 5001 bp in length
            * 118975    119074: gap of 100 bp
            * 119075    139540: contig of 20466 bp in length
            * 139541    139640: gap of 100 bp
            * 139641    146697: contig of 7057 bp in length
            * 146698    146797: gap of 100 bp
            * 146798    149204: contig of 2407 bp in length
            * 149205    149304: gap of 100 bp
            * 149305    158380: contig of 9076 bp in length
            * 158381    158480: gap of 100 bp
            * 158481    176802: contig of 18322 bp in length
            * 176803    176902: gap of 100 bp
            * 176903    179349: contig of 2447 bp in length
            * 179350    179449: gap of 100 bp
            * 179450    183727: contig of 4278 bp in length
            * 183728    183827: gap of 100 bp
            * 183828    186517: contig of 2690 bp in length
            * 186518    186617: gap of 100 bp
            * 186618    195488: contig of 8871 bp in length
            * 195489    195588: gap of 100 bp
            * 195589    214558: contig of 18970 bp in length
            * 214559    214658: gap of 100 bp
            * 214659    235423: contig of 20765 bp in length
            * 235424    235523: gap of 100 bp
            * 235524    283867: contig of 48344 bp in length
            * 283868    283967: gap of 100 bp
            * 283968    302319: contig of 18352 bp in length
            * 302320    302419: gap of 100 bp
            * 302420    309895: contig of 7476 bp in length
            * 309896    309995: gap of 100 bp
            * 309996    315745: contig of 5750 bp in length
            * 315746    315845: gap of 100 bp
            * 315846    333642: contig of 17797 bp in length
            * 333643    333742: gap of 100 bp
            * 333743    350641: contig of 16899 bp in length
            * 350642    350741: gap of 100 bp
            * 350742    365576: contig of 14835 bp in length
            * 365577    365676: gap of 100 bp
            * 365677    370201: contig of 4525 bp in length
            * 370202    370301: gap of 100 bp
            * 370302    377234: contig of 6933 bp in length
            * 377235    377334: gap of 100 bp
            * 377335    394656: contig of 17322 bp in length
            * 394657    394756: gap of 100 bp
            * 394757    410877: contig of 16121 bp in length.
FEATURES             Location/Qualifiers
     source          1..410877
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-58J8"
                     /clone_lib="DanioKey"
     misc_feature    1..5913
                     /note="assembly_fragment:00193
                     fragment_chain:1"
     misc_feature    6014..8737
                     /note="assembly_fragment:00080
                     fragment_chain:1"
     misc_feature    8838..12348
                     /note="assembly_fragment:01015
                     fragment_chain:1"
     misc_feature    12449..34866
                     /note="assembly_fragment:00049
                     fragment_chain:1"
     misc_feature    34967..50012
                     /note="assembly_fragment:00641
                     fragment_chain:1"
     misc_feature    50113..52494
                     /note="assembly_fragment:00041
                     fragment_chain:1"
     misc_feature    52595..55884
                     /note="assembly_fragment:00069
```

```
     misc_feature    fragment_chain:1"
                     55985..59774
                     /note="assembly_fragment:00091
     misc_feature    fragment_chain:1"
                     59875..71704
                     /note="assembly_fragment:00583
     misc_feature    fragment_chain:1"
                     71805..90091
                     /note="assembly_fragment:00724
     misc_feature    fragment_chain:1"
                     90192..94302
                     /note="assembly_fragment:00156
     misc_feature    fragment_chain:1"
                     94403..104620
                     /note="assembly_fragment:00527
     misc_feature    fragment_chain:1"
                     104721..110745
                     /note="assembly_fragment:00172
     misc_feature    fragment_chain:1"
                     110846..113873
                     /note="assembly_fragment:00140
     misc_feature    fragment_chain:1"
                     113974..118974
                     /note="assembly_fragment:00223
     misc_feature    fragment_chain:1"
                     119075..139540
                     /note="assembly_fragment:01515
     misc_feature    fragment_chain:1"
                     139641..146697
                     /note="assembly_fragment:00319
     misc_feature    fragment_chain:1"
                     146798..149204
                     /note="assembly_fragment:00059
     misc_feature    fragment_chain:1"
                     149305..158380
                     /note="assembly_fragment:00472
     misc_feature    fragment_chain:1"
                     158481..176802
                     /note="assembly_fragment:00908
     misc_feature    fragment_chain:1"
                     176903..179349
                     /note="assembly_fragment:00034
     misc_feature    fragment_chain:1"
                     179450..183727
                     /note="assembly_fragment:00102
     misc_feature    fragment_chain:1"
                     183828..186517
                     /note="assembly_fragment:00126
                     fragment_chain:1"

Alignment Scores:
Pred. No.:            977                Length:              110000
Score:               10.00              Matches:             10
Percent Similarity:  100.0%             Conservative:        0
Best Local Similarity: 100.0%           Mismatches:          0
Query Match:         4.7%               Indels:              0
DB:                  14                 Gaps:                0

US-09-989-890-238 (1-212) x CR555291_0 (1-110000)

Qy    27 SerLeuAlaLeuLeuSerArgProLeuSer 36
         ||||||||||||||||||||||||||||||
Db 40146 AGCCTAGCATTGTATCAAGACCGTTGAGC 40117

RESULT 45
HS29C18
LOCUS       HS29C18      143701 bp   DNA      linear  PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP1-29C18 on chromosome 22, complete
            sequence.
ACCESSION   297192
VERSION     297192.2  GI:6572206
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 143701)
AUTHORS     Burton,J.
TITLE       Direct Submission
JOURNAL     Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 13, 1999 this sequence version replaced gi:2996576.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            RP1-29C18 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1..143701
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="R2PD:RPCIP704C1829"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="RP1-29C18"
                     /clone_lib="RPCI-1"
     misc_feature    1
                     /note="clone_left_end: RP1-29C18"
     gene            complement(82051..119166)
                     /locus_tag="RP1-29C18.7-001"
     mRNA            complement(join(82051..82288,86075..86309,86486..86564,
                     119033..119166))
                     /locus_tag="RP1-29C18.7-001"
     CDS             complement(join(82106..82288,86075..86309,86486..86564,
                     119033..119087))
                     /locus_tag="RP1-29C18.7-001"
                     /standard_name="OTTHUMP00000028703"
                     /codon_start=1
                     /protein_id="CAI20086.1"
                     /db_xref="GI:56208513"
                     /db_xref="InterPro:IPR001005"
                     /db_xref="UniProt/TrEMBL:Q5R3D1"
                     /translation="MIVFPBTSQPRFLYSLQNYAVRDGLGTILRGGAEBGSYDNWPHT
                     RKSWGPLSPGHQRELWTQDPDWTEVLSGHKGDAGAGCGCFCSQFINARCAHPLCLAR
                     GLDRRASBEMPILQALCLLPKENLHDASRRVHRGRGEENVEWTGDLHPAQLGRGGRPE
                     ICHSLEAIALVTKARHRKTNTTF"
     misc_feature    143701
                     /note="clone_right_end: RP1-29C18"
ORIGIN

Alignment Scores:
Pred. No.:           1.23e+03           Length:              143701
Score:               10.00              Matches:             10
Percent Similarity:  100.0%             Conservative:        0
```

```
Best Local Similarity:   100.0%          Mismatches:      0
Query Match:               4.7%          Indels:          0
DB:                          8           Gaps:            0

US-09-989-890-238 (1-212) x HS29C18 (1-143701)

Qy       26 SerSerLeuAlaLeuSerArgProLeu 35
            |||||||||||||||||||||||||||
Db     86112 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 86141

RESULT 46
AL355132/c
LOCUS       AL355132              151261 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-181D10 on chromosome 13 Contains
            a ring finger protein 12 (RNF12) pseudogene, the 5' end of the
            FOXO1A gene for forkhead box O1A (rhabdomyosarcoma), the 3' end of
            the MRPS31 gene for mitochondrial ribosomal protein S31 and a CpG
            island, complete sequence.
ACCESSION   AL355132
VERSION     AL355132.19  GI:15717908
KEYWORDS    HTG; CpG island; FOXO1A; MRPS31; RNF12.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 151261)
AUTHORS     Tromans,A.
TITLE       Direct Submission
JOURNAL     Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 20, 2001 this sequence version replaced gi:15020887.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-181D10 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            --------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            ---------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1..151261
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-181D10"
                     /clone_lib="RPCI-11.1"
     misc_feature    1
                     /note="Clone_left_end: RP11-181D10"
     gene            join(13415..13675,14001..15179,15786..16025)
                     /locus_tag="RP11-181D10.2-001"
                     /pseudo
     CDS             join(13415..13675,14001..15179,15786..16025)
```

```
                     /locus_tag="RP11-181D10.2-001"
                     /note="match: proteins: Q91X19 Q9CYY2 Q9NVW2 Q9W677
                     Q9WTV7"
                     /pseudo
                     /codon_start=1
     gene            join(complement(60260..61274),
                     complement(AL133318.11:39120..44316))
                     /gene="FOXO1A"
     mRNA            join(complement(60260..61274),
                     complement(AL133318.11:42965..44316),
                     complement(AL133318.11:39120..42490))
                     /gene="FOXO1A"
                     /locus_tag="RP11-89L15.1-001"
                     /product="forkhead box O1A (rhabdomyosarcoma)"
                     /note="match: ESTs: AA019811 AA134749 AA653517 AA727282
                     AA746346 AA856624 AA866177 AI003947 AI188507 AI418347
                     AI475740 AI523901 AI699468 AI741811 AI934945 AI936621
                     AV646553 AV646573 AW020440 AW028783 AW121569 AW300151
                     AW316979 AW367874 AW467069 AW476550 AW593488 AW621146
                     AW949943 AW950024 BB633141 BE669435 BE850923 BF434989
                     BF467469 BF509778 BF732882 BF733004 BF957712 BI044120
                     BI183684 BI233985 BI414157 BI824376 BI838883 BM453915
                     BM475689 BM678665 BM703023 BQ006072 BQ186082 BQ230226
                     W07885
                     match: cDNAs: AF032885 AF114258 AF114261 AF126056 AF178854
                     AF384029 AI252157 AY040320 AY094061 BC021981 U02308 U02310
                     U02368 U36922"
     CDS             join(complement(60260..60889),
                     complement(AL133318.11:42979..44316))
                     /gene="FOXO1A"
                     /locus_tag="RP11-89L15.1-001"
                     /standard_name="OTTHUMP00000018301"
                     /note="match: proteins: AAH21981 Q12778 Q9JW4 Q9R1E0
                     Q9W7F9 Q9WVH5"
                     /codon_start=1
                     /product="forkhead box O1A (rhabdomyosarcoma)"
                     /protein_id="CAH70978.1"
                     /db_xref="GI:55663268"
                     /translation="MAEAPQVVEIDDPFEPLPRPRSCTWPLPRPEFSQSNSATSSPAP
                     SGSAAANPDAAAGLPSASAAAVSADFMSNLSLLEEBSEDFPQAPGSVAAAVAAAAAAA
                     TGGLCGDFQGPEAGCLHPAPPQPPPGCPLSQHPVPPAAAGPLAGQPRKSSSSRRNAW
                     GNLSYADLITKAIESSAEKRLTLSQIYEWMVKSVPYFKDKGDSNSSAGWKNSIRHNLS
                     LHSKFIRVQNEGTGKSSWWMLNPEGGKSGKSPRRRAASMDNNSKFAKSRSRAAKKKAS
                     LQSGQEGAGDSPGSQFSKWPASPGSHSNDDFDNWSTFRPRTSSNASTISGRLSPIMTE
                     QDDLGEGDVHSMVYPPSAAKMASTLPSLSEISNPENMENLLDNLNLLSSPTSLTVSTQ
                     SSPGTMWQQTPCYSFAPPNTSLNSPSPNYQKYTYGQSMSPLPQMPIQTLQDNKSSYG
                     GMSQYNCAPGLLKELITSDSPPHNDIMTPVDPGVAQPNSRVLGQNVMMGPNSVMSTYG
                     SOASHNKMMNPSSHTHPGHAQQTSAVNGRPLPHTVSTMPHTSGMNRLTQVKTPVQVPL
                     PHPMQMSALGGYSSVSSCNGYGRMGLLHQEKLPSDLDGMFIERLDCDMESIIRNDLMD
                     GDTLDFNFDNVLPNQSFPHSVKTTHSWVSG"
     misc_feature    60831..60998
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    108022
                     /note="Clone_left_end: RP11-251C7"
     gene            join(complement(AL161614.16:16400..16588),
                     complement(AL161614.16:12161..12448),
                     complement(AL161614.16:4363..4521),
                     complement(AL161614.16:2288..2428),
                     complement(148695..148768),complement(143814..143957),
                     complement(123972..124277))
                     /gene="MRPS31"
                     /locus_tag="RP11-181D10.3-001"
     mRNA            join(complement(AL161614.16:16400..16588),
                     complement(AL161614.16:12161..12448),
                     complement(AL161614.16:4363..4521),
                     complement(AL161614.16:2288..2428),
                     complement(148695..148768),complement(143814..143957),
                     complement(123972..124277))
                     /gene="MRPS31"
                     /locus_tag="RP11-181D10.3-001"
```

```
Qy          52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
               ||||||||||||||||||||||||||||||
Db       60724 GCGGGCTGCCCTCCGCCTCGGCTGCCGCT 60695

RESULT 47
AC005901/c
LOCUS       AC005901            156763 bp    DNA      linear   PRI 20-JAN-1999
DEFINITION  Homo sapiens chromosome 17, clone hRPK.15_K_2, complete sequence.
ACCESSION   AC005901
VERSION     AC005901.1  GI:415016
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 156763)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.15_K_2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156763)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,R., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., McKernan,K., Meldrim,J., Molla,M., Morris,N., McEwan,P.,
            McGurk,A., McKernan,K., Morris,W., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 156763)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Naylor,J., Niloff,M., O'Connor,T., Riley,R., Morrow,J., Mychaleckyj,J.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 20, 1999 this sequence version replaced gi:4153856.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            Only 156763 base pairs from the middle of this clone are being
            submitted. The remainder overlaps either accession number AC002994
            (WICGR project L273) or accession number AC005746 (WICGR project
            L399).
FEATURES             Location/Qualifiers
     source          1..156763
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
```

```
                     /product="mitochondrial ribosomal protein S31"
                     /note="match: ESTs: AA742194 AA886331 AA947589 AI077473
                     AI127882 AI343737 AI344511 AI636049 AI700828 AI741536
                     AL518731 AL518732 AL525614 AW083272 AW629020 BE559706
                     BF111436 BF690564 BG499349 BG614293 BG615490 BG615614
                     BG615812 BG616392 BG722006 BI185560 BI821428 BM980164
                     BQ006817
                     match: cDNAs: BC010159 BC022045 Z68747"
     gene            complement(join(123972. .124277,143814. .144289))
                     /gene="MRPS31"
                     /locus_tag="RP11-181D10.3-004"
     mRNA            complement(join(123972. .124277,143814. .144289))
                     /gene="MRPS31"
                     /locus_tag="RP11-181D10.3-004"
                     /product="mitochondrial ribosomal protein S31"
                     /note="match: ESTs: AA461355 AW274392 BQ002789 W79354"
     polyA_site      complement(123972)
                     /gene="MRPS31"
     polyA_signal    complement(123993. .123998)
                     /gene="MRPS31"
     CDS             join(complement(AL161614.16:16400. .16551),
                     complement(AL161614.16:12161. .12448),
                     complement(AL161614.16:4363. .4521),
                     complement(AL161614.16:2288. .2428),
                     complement(148695. .148768),complement(143814. .143957),
                     complement(124048. .124277))
                     /locus_tag="RP11-181D10.3-004"
                     /gene="MRPS31"
                     /locus_tag="RP11-181D10.3-001"
                     /standard_name="OTTHUMP00000018305"
                     /note="match: proteins: Q61733 Q8VEH4 Q8WTV8 Q92665 Q961X2
                     Q9D045"
                     /codon_start=1
                     /product="mitochondrial ribosomal protein S31"
                     /protein_id="CAH70977.1"
                     /db_xref="GI:55663267"
                     /db_xref="GOA:Q5VYC8"
                     /db_xref="UniProt/TrEMBL:Q5VYC8"
                     /translation="MFPRVSTFLPLRPLSRHPLSSGSPETSAAAIMLLTVRHGTVRYR
                     SSALLARTKNNIQRYFGTNSVICSKKDKQSVRTEETSKETSESSQDSEKENTKKDLLGI
                     IKGMKVRLSTVNVRTRTKPPKRPLKSLEATLGRLRRATEYAPKKRIEPLSPELVAAAS
                     AVADSLPFDKQTTKSELLSQLQQHEESRAQRDAKRPKISFSNIISDMKVARSATARV
                     RSRPELRIQFDEGYDNYPGQEKTDDLKKRKNIFTGKRLNIFDMMAVTKEAPETDTSPS
                     LWDVEFAKQLATVNEQPLQNGFEELIQWTKEGKLWEFPINNEAGFDDDGSEFHEHIFL
                     EKHLESFPPKQGPIRHFMELVTCGLSKNPYLSVKQKVEHIEWFRNYFNEKKDILKESNI
                     QFN"
     gene            complement(join(124050. .124277,143814. .143957,
                     146949. .147083))
                     /gene="MRPS31"
     mRNA            complement(join(124050. .124277,143814. .143957,
                     146949. .147083))
                     /gene="MRPS31"
                     /locus_tag="RP11-181D10.3-003"
                     /product="mitochondrial ribosomal protein S31"
                     /note="match: cDNAs: BM142076"
     misc_feature    133024
                     /note="Clone_right_end: RP11-14H2"
     misc_feature    151261
                     /note="Clone_right_end: RP11-181D10"
ORIGIN
```

```
Alignment Scores:
Pred. No.:            1.28e+03      Length:           151261
Score:              10.00          Matches:          10
Percent Similarity: 100.0%         Conservative:     0
Best Local Similarity: 100.0%      Mismatches:       0
Query Match:         4.7%          Indels:           0
DB:                  8            Gaps:             0

US-09-989-890-238 (1-212) x AL355132 (1-151261)
```

```
                    /map="17"
                    /clone="hRPK.15_K_2"
                    /clone_lib="RPCI-11 human BAC library"
repeat_region       1..59
                    /rpt_family="AluJ/FRAM"
repeat_region       60..79
                    /rpt_family="(CAAA)n"
repeat_region       1165..1448
                    /rpt_family="AluJo"
repeat_region       1449..1466
                    /rpt_family="AT_rich"
repeat_region       2792..2846
                    /rpt_family="AT_rich"
repeat_region       complement(2847..3643)
                    /rpt_family="L1PA2"
repeat_region       complement(3664..3770)
                    /rpt_family="L2"
repeat_region       3868..3894
                    /rpt_family="AT_rich"
repeat_region       complement(3946..4280)
                    /rpt_family="L1PA8"
repeat_region       4381..4841
                    /rpt_family="L2"
repeat_region       4870..5160
                    /rpt_family="AluSq"
repeat_region       complement(5364..5679)
                    /rpt_family="AluSx"
repeat_region       5752..5897
                    /rpt_family="L1MC2"
repeat_region       5976..6441
                    /rpt_family="L1MC2"
repeat_region       6485..6765
                    /rpt_family="AluSx"
repeat_region       6766..6792
                    /rpt_family="(CAAA)n"
repeat_region       6898..7181
                    /rpt_family="AluJo"
repeat_region       7220..7252
                    /rpt_family="(CAAA)n"
repeat_region       complement(8690..8991)
                    /rpt_family="AluY"
repeat_region       9430..9748
                    /rpt_family="AluJb"
repeat_region       10170..10465
                    /rpt_family="AluJb"
repeat_region       complement(10752..11055)
                    /rpt_family="AluYb8"
repeat_region       complement(11613..11702)
                    /rpt_family="AluSx"
repeat_region       11873..12179
                    /rpt_family="AluSx"
repeat_region       complement(12182..12244)
                    /rpt_family="Alu"
repeat_region       13394..13435
                    /rpt_family="AT_rich"
repeat_region       13780..13800
                    /rpt_family="AT_rich"
repeat_region       14820..15103
                    /rpt_family="AluSx"
repeat_region       15104..15129
                    /rpt_family="MIR"
repeat_region       15156..15186
                    /rpt_family="(CA)n"
repeat_region       complement(15736..16045)
                    /rpt_family="AluSx"
repeat_region       complement(16701..16813)
                    /rpt_family="MIR"
repeat_region       16929..17231
                    /rpt_family="AluSg1"
repeat_region       17328..17614
                    /rpt_family="AluSc"
repeat_region       17615..17920
                    /rpt_family="AluSc"

repeat_region       17921..18041
                    /rpt_family="GA-rich"
repeat_region       18870..19391
                    /rpt_family="L2"
repeat_region       19632..19829
                    /rpt_family="MER20"
repeat_region       19865..19936
                    /rpt_family="L2"
repeat_region       20206..20319
                    /rpt_family="HAL1"
repeat_region       20402..20568
                    /rpt_family="HAL1"
repeat_region       20914..21001
                    /rpt_family="A-rich"
repeat_region       21044..22134
                    /rpt_family="L1M4"
unsure              complement(21730..21766)
                    /note="Single-stranded terminator coverage."
unsure              21882..21885
                    /note="Three subclones have four T's here; one subclone
                    has five T's here."
repeat_region       22416..23364
                    /rpt_family="L1M4"
repeat_region       23364..25592
                    /rpt_family="L1MC2"
repeat_region       25593..25861
                    /rpt_family="L1MC2"
repeat_region       25864..26157
                    /rpt_family="AluJb"
repeat_region       26158..26379
                    /rpt_family="AluY"
repeat_region       complement(26380..26676)
                    /rpt_family="L1MC2"
repeat_region       26677..27262
                    /rpt_family="AluSq"
repeat_region       complement(27263..27464)
                    /rpt_family="L1MC2"
repeat_region       27465..27738
                    /rpt_family="AluSx"
repeat_region       complement(27739..27786)
                    /rpt_family="Tigger1"
repeat_region       complement(27780..28199)
                    /rpt_family="Tigger1"
repeat_region       28207..28274
                    /rpt_family="Tigger1"
repeat_region       28275..28316
                    /rpt_family="AluSx"
repeat_region       complement(28317..30059)
                    /rpt_family="L1PA16"
repeat_region       30080..30372
                    /rpt_family="(CA)n"
repeat_region       complement(30373..30516)
                    /rpt_family="L1PA16"
repeat_region       30517..30816
                    /rpt_family="AluSq"
repeat_region       complement(30817..31281)
                    /rpt_family="Tigger1"
repeat_region       complement(31282..31557)
                    /rpt_family="Tigger1"
repeat_region       complement(31558..31622)
                    /rpt_family="AluSx"


Alignment Scores:
Pred. No.:
Score:                  1.32e+03        Length:         156763
Percent Similarity:     10.00           Matches:        10
Best Local Similarity:  100.0%          Conservative:   0
Query Match:            100.0%          Mismatches:     0
DB:                     4.7%            Indels:         0
                        8               Gaps:           0

US-09-989-890-238 (1-212) x AC005901 (1-156763)

Qy    23 AlaLeuProSerSerLeuAlaLeuLeuSer 32
         ||||||||||||||||||||||||||||||
```

Db   113671 GCCCTGCCCTCCAGCCTGGCTCTCCTTTCC 113642

RESULT 48
CR936217/c
LOCUS   CR936217   174869 bp   DNA   linear   HTG 23-JUN-2005
DEFINITION   Danio rerio chromosome 24 clone DKEYP-122B1, WORKING DRAFT
                SEQUENCE, 4 unordered pieces.
ACCESSION   CR936217
VERSION   CR936217.2 GI:68162272
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Danio rerio (zebrafish)
ORGANISM   Danio rerio
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
                Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 174869)
AUTHORS   McLaren,S.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
                Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                zfish-help@sanger.ac.uk Clone requests:
                http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT   On Jun 23, 2005 this sequence version replaced gi:58418249.
                --------- Genome Center
                Center: Wellcome Trust Sanger Institute
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: zfish-help@sanger.ac.uk
                --------- Project Information
                Center project name: zKp122B1
                --------- Summary Statistics
                Assembly program: XGAP4; version 4.5
                Chemistry: Dye-terminator; 100% of reads
                Consensus quality: 173726 bases at least Q40
                Consensus quality: 173930 bases at least Q30
                Consensus quality: 174104 bases at least Q20
                Insert size: 174569; sum-of-contigs
                Insert size: 169179; 7.2% error; agarose-fp
                Quality coverage: 9.10x in Q20 bases; sum-of-contigs Quality
                coverage: 9.61x in Q20 bases; agarose-fp
                ----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 4 contigs. The true order of the pieces
                * is not known and their order in this sequence record as
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *   1   96348: contig of 96348 bp in length
                *   96348   96448: gap of 100 bp
                *   96449   153731: contig of 57283 bp in length
                *   153732   153831: gap of 100 bp
                *   153832   157626: contig of 3795 bp in length
                *   157627   157726: gap of 100 bp
                *   157727   174869: contig of 17143 bp in length.
FEATURES   Location/Qualifiers
     source   1..174869
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="24"
                     /clone="DKEYP-122B1"
                     /clone_lib="DanioKeypilot"
     misc_feature   1..96348
                     /note="assembly_fragment:01029
                     fragment_chain:1"
     misc_feature   96449..153731
                     /note="assembly_fragment:00281
                     fragment_chain:1"
     misc_feature   153832..157626
                     /note="assembly_fragment:00001.0"
     misc_feature   157727..174869

ORIGIN   /note="assembly_fragment:00052"

Alignment Scores:
Pred. No.:   1.45e+03   Length: 174869
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity:100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x CR936217 (1-174869)

Qy   31 LeuSerArgProLeuSerProProProAla 40
             ||||||||||||||||||||||||||||||
Db   52087 CTCTCCAGACCACTAAGTCCGCCGCCTGCG 52058

RESULT 49
CR846094
LOCUS   CR846094   175626 bp   DNA   linear   VRT 26-MAY-2005
DEFINITION   Zebrafish DNA sequence from clone CH211-264F16 in linkage group 24,
                complete sequence.
ACCESSION   CR846094
VERSION   CR846094.7 GI:6392950
KEYWORDS   HTG.
SOURCE   Danio rerio (zebrafish)
ORGANISM   Danio rerio
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
                Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 175626)
AUTHORS   Dunn,M.
TITLE   Direct Submission
JOURNAL   Submitted (26-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
                Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On May 22, 2005 this sequence version replaced gi:66351741.
                --------- Genome Center
                Center: Wellcome Trust Sanger Institute
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: zfish-help@sanger.ac.uk

                During sequence assembly data is compared from overlapping clones.
                Where differences are found these are annotated as variations
                together with a note of the overlapping clone name. Note that the
                variation annotation may not be found in the sequence submission
                corresponding to the overlapping clone, as we submit sequences with
                only a small overlap as described above.
                This sequence was finished as follows unless otherwise noted: all
                regions were either double-stranded or sequenced with an alternate
                chemistry or covered by high quality data (i.e. phred quality >=
                30); an attempt was made to resolve all sequencing problems, such
                as compressions and repeats; all regions were covered by at least
                one plasmid subclone or more than one M13 subclone; and the
                assembly was confirmed by restriction digest, except on the rare
                occasion of the clone being a YAC.
                The following abbreviations are used to associate primary accession
                numbers given in the feature table with their source databases:
                Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
                on the WORMPEP database can be found at
                http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
                Zebrafish pUC subclones occasionally display inconsistency over the
                length of mononucleotide A/T runs and conserved TA repeats. Where
                this is found the longest good quality representation will be
                submitted.
                Repeat names beginning 'Dr' were identified by the Recon repeat
                discovery system (Zhirong Bao and Sean Eddy, submitted), and those
                beginning 'drr' were identified by Rick Waterman (Stephen Johnson
                lab, WashU). For further information see
                http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
                CH211-264F16 is from a CHORI-211 BAC library
                VECTOR: pTARBAC2.1.

```
FEATURES             Location/Qualifiers
     source          1..175626
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-264F16"
                     /clone_lib="CHORI-211"
ORIGIN

Alignment Scores:
Pred. No.:              1.46e+03        Length:              175626
Score:                  10.00           Matches:             10
Percent Similarity:     100.0%          Conservative:        0
Best Local Similarity:  100.0%          Mismatches:          0
Query Match:            4.7%            Indels:              0
DB:                                     Gaps:                5

US-09-989-890-238 (1-212) x CR846094 (1-175626)

Qy      31 LeuSerArgProLeuSerProProProAla 40
           |||||||||||||||||||||||||||||||
Db  128819 CTCTCCAGACCACTAAGTCCGCCGCCTGCG 128848

RESULT 50
AC137538/c
LOCUS       AC137538      199160 bp    DNA     linear   HTG 18-DEC-2002
DEFINITION  Canis familiaris clone RP81-349K11, WORKING DRAFT SEQUENCE, 2
            ordered pieces.
ACCESSION   AC137538
VERSION     AC137538.2  GI:27228818
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 199160)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paquirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripp,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199160)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 199160)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Dec 18, 2002 this sequence version replaced gi:25167121.
            -------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
            Center project name: dgr
            Center clone name: 349K11
            --------------
            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 198828 bases at least Q40
            Consensus quality: 198947 bases at least Q30
            Consensus quality: 199018 bases at least Q20
            Insert size: 185000; agarose-fp
            Insert size: 199060; sum-of-contigs
            Quality coverage: 13.82x in Q20 bases; agarose-fp
            Quality coverage: 12.84x in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   117196: contig of 117196 bp in length
            *  117197   117296: gap of unknown length
            *  117297   199160: contig of 81864 bp in length.

FEATURES             Location/Qualifiers
     source          1..199160
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9615"
                     /clone="RP81-349K11"
                     /clone_lib="RP81"
     misc_feature    1..117196
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             117197..117296
                     /estimated_length=unknown
     misc_feature    117297..199160
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

Alignment Scores:
Pred. No.:              1.63e+03        Length:              199160
Score:                  10.00           Matches:             10
Percent Similarity:     100.0%          Conservative:        0
Best Local Similarity:  100.0%          Mismatches:          0
Query Match:            4.7%            Indels:              0
DB:                                     Gaps:                14

US-09-989-890-238 (1-212) x AC137538 (1-199160)

Qy      23 AlaLeuProSerSerLeuAlaLeuLeuSer 32
           ||||||||||||||||||||||||||||||
Db  194306 GCTCTCCCAGCAGTCTTGCCCTGCTGAGT 194277

RESULT 51
AC124555/c
LOCUS       AC124555      210174 bp    DNA     linear   ROD 01-JAN-2004
DEFINITION  Mus musculus BAC clone RP23-245K17 from chromosome 14, complete
            sequence.
ACCESSION   AC124555
VERSION     AC124555.4  GI:38229416
KEYWORDS    HTG.
```

```
SOURCE        Mus musculus (house mouse)
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 210174)
  AUTHORS     Kruchowski,S. and Haakenson,W.
  TITLE       The sequence of Mus musculus BAC clone RP23-245K17
  JOURNAL     Unpublished (2001)
REFERENCE     2 (bases 1 to 210174)
  AUTHORS     Wilson,R.
  TITLE       Sequencing of Mus musculus
  JOURNAL     Unpublished (2001)
REFERENCE     3 (bases 1 to 210174)
  AUTHORS     McPherson,J.D. and Waterston,R.H.
  TITLE       Direct Submission
  JOURNAL     Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
REFERENCE     4 (bases 1 to 210174)
  AUTHORS     Wilson,R.K.
  TITLE       Direct Submission
  JOURNAL     Submitted (20-JUN-2003) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
REFERENCE     5 (bases 1 to 210174)
  AUTHORS     Wilson,R.K.
  TITLE       Direct Submission
  JOURNAL     Submitted (08-NOV-2003) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
REFERENCE     6 (bases 1 to 210174)
  AUTHORS     Wilson,R.
  TITLE       Direct Submission
  JOURNAL     Submitted (01-JAN-2004) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT       On Nov 8, 2003 this sequence version replaced gi:31880269.
              --------------- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu
              Contact: submissions@watson.wustl.edu
              --------------- Summary Statistics
              Center project name: M_BA0245K17
              ---------------.

              NOTICE:  This sequence may not represent the entire insert of this
              clone.  It may be shorter because we only sequence overlapping
              clone sections once, or longer because we provide a small overlap
              between neighboring data submissions.

              This sequence was finished as follows unless otherwise noted:
              all regions were double stranded, sequenced with an alternate
              chemistry, or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by sequence
              from more than one subclone; and the assembly was confirmed by
              restriction digest.

              MAPPING INFORMATION:
              Mapping information for this clone was provided by Dr. Wes Warren,
              Department of Genetics, Washington University, St. Louis MO. For
              additional information about the map position of this sequence, see
              http://genome.wustl.edu

              SOURCE INFORMATION:
              The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
              and Minako Tateno in the laboratory of Pieter de Jong
              (http://www.chori.org) from female C57BL/6J mouse kidney and/or
              brain genomic DNA.  The clone and detailed information can be
              obtained from Research Genetics, Inc. (http://www.resgen.com) or
              Pieter de Jong and coworkers at http://www.chori.org

              NEIGHBORING SEQUENCE INFORMATION:
              This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..210174
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="14"
                     /map="14"
                     /clone="RP23-245K17"
                     /clone_lib="RPCI-23"
     repeat_region   1820..1922
                     /rpt_family="Alu"
     repeat_region   1924..2112
                     /rpt_family="L1"
     repeat_region   2376..2919
                     /rpt_family="ERV1"
     repeat_region   3164..3207
                     /rpt_family="ID"
     repeat_region   4294..4552
                     /rpt_family="L1"
     repeat_region   6511..6676
                     /rpt_family="B4"
     repeat_region   7565..7625
                     /rpt_family="ID"
     misc_feature    7958..8813
                     /note="CpG_island (%GC=70.9, o/e=0.79, #CpGs=92)"
     repeat_region   10008..10146
                     /rpt_family="L1"
     repeat_region   10158..10240
                     /rpt_family="Alu"
     repeat_region   10591..10737
                     /rpt_family="Alu"
     repeat_region   10812..11087
                     /rpt_family="B4"
     repeat_region   12120..12350
                     /rpt_family="MaLR"
     repeat_region   12351..12739
                     /rpt_family="MaLR"
     repeat_region   12724..13063
                     /rpt_family="MaLR"
     repeat_region   13064..13127
                     /rpt_family="MaLR"
     repeat_region   13128..13212
                     /rpt_family="MaLR"
     repeat_region   13939..14339
                     /rpt_family="L1"
     misc_feature    14331..14508
                     /note="Sequence derived from PCR product of project DNA."
     repeat_region   14667..14721
                     /rpt_family="MaLR"
     repeat_region   14748..15064
                     /rpt_family="MaLR"
     repeat_region   15214..15545
                     /rpt_family="ID"
     repeat_region   16682..17152
                     /rpt_family="MaLR"
     repeat_region   17359..17656
                     /rpt_family="MaLR"
     repeat_region   18207..18276
                     /rpt_family="MaLR"
     repeat_region   18446..18763
                     /rpt_family="L1"
     repeat_region   20646..20709
                     /rpt_family="ERV1"
     repeat_region   21743..21980
                     /rpt_family="ERVK"
     repeat_region   22601..22673
                     /rpt_family="ID"
     repeat_region   23916..24235
                     /rpt_family="MaLR"
     repeat_region   24257..24382
                     /rpt_family="L1"
     repeat_region   24401..24541
                     /rpt_family="Alu"
     repeat_region   24577..24717
```

```
repeat_region      /rpt_family="L1"       25004..25313
repeat_region      /rpt_family="MaLR"     27461..27567
repeat_region      /rpt_family="L2"       27632..27748
repeat_region      /rpt_family="L2"       28055..28437
repeat_region      /rpt_family="ERV1"     28485..28593
repeat_region      /rpt_family="ERV1"     29303..29595
repeat_region      /rpt_family="ERV1"     29596..29794
repeat_region      /rpt_family="B2"       29795..30347
repeat_region      /rpt_family="ERV1"     30364..30598
repeat_region      /rpt_family="B4"       30599..30798
repeat_region      /rpt_family="B2"       30795..30869
repeat_region      /rpt_family="B4"       30933..31050
repeat_region      /rpt_family="B2"       31027..31140
repeat_region      /rpt_family="B2"       31177..31432
repeat_region      /rpt_family="ERV1"     32042..32109
repeat_region      /rpt_family="ERV1"     32777..32923
repeat_region      /rpt_family="Alu"      33027..33114
repeat_region      /rpt_family="B2"       33274..33381
repeat_region      /rpt_family="B4"       34425..34571
repeat_region      /rpt_family="Alu"      35277..35353
repeat_region      /rpt_family="ID"       35502..35933
repeat_region      /rpt_family="ERVK"     36442..36671
repeat_region      /rpt_family="MaLR"     39264..39307
repeat_region      /rpt_family="L1"       39303..39362
```

```
Alignment Scores:
Pred. No.:                          Length:        210174
Score:                  1.7e+03     Matches:       10
Percent Similarity:     10.00       Conservative:  0
Best Local Similarity:  100.0%      Mismatches:    0
Query Match:            100.0%      Indels:        0
DB:                     4.7%        Gaps:          0
                        9

US-09-989-890-238 (1-212) x AC124555 (1-210174)

Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||||
Db  134995 GCTGGGGCTGCCTAGTGCCAGCGCTGCTGCG 134966

RESULT 52
AC158593
LOCUS       AC158593    216621 bp   DNA    linear   ROD 30-MAY-2005
DEFINITION  Mus musculus 10 BAC RP23-197F5 (Roswell Park Cancer Institute
            (C57BL/6J Female) Mouse BAC Library) complete sequence.
ACCESSION   AC158593
VERSION     AC158593.8  GI:66793646
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 216621)
  AUTHORS   Muzny,D., Adams,C., Agbai II,O., Allen,C., Alsbrooks,S., Archer,P.,
            Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
            Beraducci,A., Biswalo,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
            Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
            Chacko,J., Chahrour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
            Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,
            Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
            Anda,C., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H.,
            Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
            Dziuda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
            Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
            Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
            Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,
            Gonzalez-Garay,M., Guevara,W., Holder,M., Haaland,W., Haeberlen,K.,
            Hall,B., Hamid,H., Hamilton,K., Harris,R., Havlak,P.,
            Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
            Hines,S., Hitchens,M., Hodgson,A., Hogues,M., Hollins,B.,
            Howell,L.T., Hulyk,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,
            Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S.,
            Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,F.,
            Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B.,
            Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyanage,D.,
            London,P., Lopez,J., Lorensuhewa,L., Lozado,R., Luk,T., Madu,R.,
            Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
            McClelland,H., McPherson,J., Mercadao,C., Metzker,M.,
            Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Munidasa,M.,
            Murray,D., Nazarith,L., Ngo,D., Nguyen,N., Norwig-Eastaugh,E.,
            Nott,A., Nwaokelemeh,O., Obregon,M., Ochi-Okorie,C., Odeh,E.,
            Okwuonu,G., Okwuonu,K., Parker,D., Pasternak,S., Patel,B.,
            Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T.,
            Primus,E., Pu,L.-L., Puazo,M., Qin,X., Quinn,A., Quiroz,J.,
            Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,
            Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,
            Santibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,
            Sisson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R.,
            Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,
            Trejos,Z., Usmani,K., Vargo,C., Verduzco,D., Villasana,D., Virk,D.,
            Volkov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,
            Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,
            Wu,J., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L.,
            Zhang,Z., Zhou,J., Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216621)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2005) Human Genome Sequencing Center, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 216621)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2005) Human Genome Sequencing Center, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 216621)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2005) Human Genome Sequencing Center, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 30, 2005 this sequence version replaced gi:6325308.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
```

Zhang.

Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
------------ Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Drafting Center Code: WIBR
          Contact: hgsc-help@bcm.tmc.edu.

FEATURES   Location/Qualifiers
source   1..21621
                 /organism="Mus musculus"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:10090"
                 /chromosome="10"
                 /clone="RP23-197F5"
misc_feature   1..15082
                 /note="overlaps bases 146386..161467 of clone AC155323"
                 /function="clone overlap"
repeat_region   735..1122
                 /rpt_family="MTA_Mm"
repeat_region   complement(1436..1967)
                 /rpt_family="RLTR42-int"
repeat_region   complement(2027..2136)
                 /rpt_family="PB1D9"
repeat_region   2249..2295
                 /rpt_family="(TG)n"
repeat_region   complement(2296..2354)
                 /rpt_family="B1_Mm"
repeat_region   2357..2722
                 /rpt_family="ORR1B1"
repeat_region   complement(2719..2786)
                 /rpt_family="B1_Mus1"
repeat_region   complement(2787..2862)
                 /rpt_family="RLTR42-int"
repeat_region   3464..3613
                 /rpt_family="B1_Mus1"
repeat_region   3615..3647
                 /rpt_family="(CAAAAA)n"
repeat_region   complement(3705..4352)
                 /rpt_family="L1_Mus1"
repeat_region   4518..4544
                 /rpt_family="AT_rich"
repeat_region   complement(4694..4950)
                 /rpt_family="RLTR42-int"
repeat_region   complement(4969..5152)
                 /rpt_family="B3"
repeat_region   complement(5376..5480)
                 /rpt_family="RMER12"
repeat_region   complement(5470..5594)
                 /rpt_family="B1_Mm"
repeat_region   complement(5597..6015)
                 /rpt_family="RMER12"
repeat_region   6016..6152
                 /rpt_family="B1_Mus2"
repeat_region   6153..6192
                 /rpt_family="(CAAA)n"
repeat_region   6193..6203
                 /rpt_family="(CAA)n"
repeat_region   complement(6204..6506)
                 /rpt_family="RMER12"
repeat_region   6531..6778
                 /rpt_family="RLTR18B"
repeat_region   6779..6842
                 /rpt_family="L2"
repeat_region   6843..7014
                 /rpt_family="RLTR18B"
repeat_region   7017..7065
                 /rpt_family="(CA)n"
repeat_region   complement(7377..8467)
                 /rpt_family="L1Md_F2"

repeat_region   8468..8805
                 /rpt_family="L1Md_F2"
repeat_region   9709..9745
                 /rpt_family="(TTTA)n"
repeat_region   9806..9846
                 /rpt_family="(TTTA)n"
repeat_region   9907..9947
                 /rpt_family="(TTTA)n"
repeat_region   complement(9999..10194)
                 /rpt_family="Tigger7"
repeat_region   complement(10230..10968)
                 /rpt_family="RMER1B"
repeat_region   complement(11013..11211)
                 /rpt_family="ORR1A1"
repeat_region   11535..11585
                 /rpt_family="(GA)n"
repeat_region   11628..11752
                 /rpt_family="L1MD2"
repeat_region   11792..11896
                 /rpt_family="RMER15"
repeat_region   complement(11933..12025)
                 /rpt_family="L1M4"
repeat_region   12115..12186
                 /rpt_family="(TTA)n"
repeat_region   complement(12187..12497)
                 /rpt_family="Lx4A"
repeat_region   complement(12498..13439)
                 /rpt_family="Lx4A"
repeat_region   complement(13440..13576)
                 /rpt_family="B1_Mus1"
repeat_region   complement(13577..13895)
                 /rpt_family="Lx4A"
repeat_region   13919..13942
                 /rpt_family="(TTTTG)n"
repeat_region   complement(13945..14079)
                 /rpt_family="B1_Mus1"
repeat_region   14091..14114
                 /rpt_family="(T)n"
repeat_region   complement(14134..14215)
                 /rpt_family="B3"
repeat_region   complement(14180..14432)
                 /rpt_family="B4A"
repeat_region   14434..14490
                 /rpt_family="C-rich"
repeat_region   complement(14491..14610)
                 /rpt_family="B1_Mus2"

Alignment Scores:
Pred. No.:   1.75e+03   Length: 216621
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 9   Gaps: 0

US-09-989-890-238 (1-212) x AC158593 (1-216621)

Qy   46 ProGlyCysGlySerGlyAlaGlyLeuPro 55
            |||||||||||||||||||||||||||||
Db   121675 CCTGGGTGTGGCAGTGGTGCAGGGCTTCCT 121704

RESULT 53
CR792443
LOCUS   CR792443   217987 bp   DNA   linear   VRT 10-AUG-2005
DEFINITION   Zebrafish DNA sequence from clone DKEY-52K1 in linkage group 13,
             complete sequence.
ACCESSION   CR792443
VERSION   CR792443.12 GI:72419832
KEYWORDS   HTG.
SOURCE   Danio rerio (zebrafish)
  ORGANISM   Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.
```
REFERENCE   1  (bases 1 to 217987)
  AUTHORS   Wallis,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
            On Aug 12, 2005 this sequence version replaced gi:72060589.
  COMMENT   ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e. phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep
            Clone-derived Zebrafish pUC subclones occasionally display
            inconsistency over the length of mononucleotide A/T runs and
            conserved TA repeats. Where this is found the longest good quality
            representation will be submitted.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies. The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.
            DKEY-52K1 is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1..217987
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="13"
                     /clone="DKEY-52K1"
                     /clone_lib="DanioKey"
ORIGIN

Alignment Scores:
Pred. No.:              1.76e+03     Length:         217987
Score:                     10.00     Matches:            10
Percent Similarity:       100.0%     Conservative:        0
Best Local Similarity:    100.0%     Mismatches:          0
Query Match:                4.7%     Indels:              0
DB:                            5     Gaps:                0

US-09-989-890-238 (1-212) x CR792443 (1-217987)

Qy     27 SerLeuAlaLeuSerArgProLeuSer 36
          |||||||||||||||||||||||||||
Db 169805 AGCCTAGCATTGTTATCAAGACCGTTGAGC 169834

RESULT 54
AC078930
LOCUS       AC078930   226168 bp   DNA   linear   ROD 01-JAN-2001
DEFINITION  Mus musculus 10 BAC 280B6 (Roswell Park Cancer Institute Human BAC
            Library) complete sequence.
ACCESSION   AC078930
VERSION     AC078930.13  GI:12000469
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 226168)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Louiseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226168)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 226168)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 1, 2001 this sequence version replaced gi:11875267.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1..226168
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /clone="280B6"
     repeat_region   446..552
                     /rpt_family="PB1D7"
     repeat_region   684..843
                     /rpt_family="B4"
     misc_feature    complement(1333..1931)
                     /note="Region similar to Mm#S966036 L0050G07-3 Mus
                     musculus cDNA, 3' end: gb=AW551584"
     repeat_region   2083..2299
                     /rpt_family="RMER15"
     repeat_region   3610..3997
                     /rpt_family="MTA"
     repeat_region   4882..4907
                     /rpt_family="(GTTTG)n"
     repeat_region   complement(4908..5010)
                     /rpt_family="PB1D9"
     repeat_region   5011..5017
```

```
repeat_region      /rpt_family="(GTTTG)n"
                   5124..5170
repeat_region      /rpt_family="(TG)n"
                   complement(5171..5216)
repeat_region      /rpt_family="B1_MM"
                   5232..5597
repeat_region      /rpt_family="ORR1B"
                   6340..6487
repeat_region      /rpt_family="B1_MM"
                   6490..6522
repeat_region      /rpt_family="(CAAAAA)n"
                   complement(6568..7220)
repeat_region      /rpt_family="L1_MM"
                   7393..7419
repeat_region      /rpt_family="AT_rich"
                   7787..7810
repeat_region      /rpt_family="AT_rich"
                   complement(7844..8027)
repeat_region      /rpt_family="B3"
                   8253..8288
repeat_region      /rpt_family="(CA)n"
                   complement(8344..8465)
repeat_region      /rpt_family="B1_MM"
                   complement(8468..8888)
repeat_region      /rpt_family="RMER12"
                   8891..9027
repeat_region      /rpt_family="B1_MM"
                   9028..9067
repeat_region      /rpt_family="(CAAA)n"
                   9068..9081
repeat_region      /rpt_family="B1_MM"
                   complement(9074..9381)
repeat_region      /rpt_family="RMER12"
                   9892..9940
repeat_region      /rpt_family="(CA)n"
                   complement(10247..11342)
repeat_region      /rpt_family="L1_MM"
                   11342..11680
repeat_region      /rpt_family="L1"
                   12317..12380
repeat_region      /rpt_family="L2"
                   12584..12620
repeat_region      /rpt_family="RMER12"
                   12681..12721
repeat_region      /rpt_family="(TTTA)n"
                   12782..12822
repeat_region      /rpt_family="(TTTA)n"
                   complement(13130..13843)
repeat_region      /rpt_family="RMER1B"
                   complement(13888..14086)
repeat_region      /rpt_family="ORR1A1"
                   14410..14460
repeat_region      /rpt_family="(GA)n"
                   complement(14667..14771)
repeat_region      /rpt_family="RMER15"
                   complement(14808..14899)
repeat_region      /rpt_family="L1M4"
                   complement(14979..14989)
repeat_region      /rpt_family="Lx4"
                   14990..15061
repeat_region      /rpt_family="(TTA)n"
                   complement(15062..15372)
repeat_region      /rpt_family="Lx4"
                   complement(15369..16281)
repeat_region      /rpt_family="Lx3"
                   16284..16321
repeat_region      /rpt_family="(TTTG)n"
                   complement(16322..16450)
repeat_region      /rpt_family="B1_MM"
                   complement(16451..16770)
repeat_region      /rpt_family="Lx3"
                   16794..16817
repeat_region      /rpt_family="(TTTTG)n"

repeat_region      complement(16820..16954)
                   /rpt_family="B1_MM"
repeat_region      16966..16989
                   /rpt_family="(T)n"
repeat_region      complement(17055..17307)
                   /rpt_family="B4A"
repeat_region      17305..17365
                   /rpt_family="C-rich"
repeat_region      complement(17366..17485)
                   /rpt_family="B1_MM"
repeat_region      17488..17531
                   /rpt_family="(CCA)n"
repeat_region      18519..18569
                   /rpt_family="RMER17C"
repeat_region      19296..19442
                   /rpt_family="CT-rich"
repeat_region      20083..20228
                   /rpt_family="B1_MM"
repeat_region      20231..20253
                   /rpt_family="(CAAAAA)n"
repeat_region      20639..20665
                   /rpt_family="(TTTA)n"
repeat_region      complement(20666..20823)
                   /rpt_family="B2"
repeat_region      21501..21897
                   /rpt_family="MTA"
repeat_region      22097..22194
                   /rpt_family="(TC)n"
repeat_region      22217..22249
                   /rpt_family="CT-rich"
repeat_region      complement(22442..22759)
                   /rpt_family="RMER5"
repeat_region      23337..23382
                   /rpt_family="(TG)n"
repeat_region      complement(23430..23502)
                   /rpt_family="B3A"
gene               join(24188..24291,29229..29298,30575..30666,31819..31965,
                   32464..33054,33190..34840,35033,37761..37830,
                   39079..40109,40257,41072..41499,41664..41756,
                   42311..42689,42828,42909..43038,44362..44541,
                   44635..44720,52489..52586,53402..54015,54020..55022,
                   55035..58066,60279..60446,67914..68005,69656..69931,
                   70048..70384,70913..71218,71290..71419)
                   /gene="gi|924266|dbj|D42063|HUMRANBP2 Human mRNA for
                   RanBP2 (Ran-binding protein 2)"
repeat_region      24525..24545
                   /rpt_family="GC_rich"


Alignment Scores:
Pred. No.:                1.81e+03      Length:         226168
Score:                    10.00         Matches:        10
Percent Similarity:       100.0%        Conservative:   0
Best Local Similarity:    100.0%        Mismatches:     0
Query Match:              4.7%          Indels:         0
DB:                       9             Gaps:           0

US-09-989-890-238 (1-212) x AC078930 (1-226168)

Qy     46 ProGlyCysGlySerGlyAlaGlyLeuPro 55
          ||||||||||||||||||||||||||||||
Db 124548 CCTGGGTGTGGCAGTGGTGCAGGGCTTCCT 124577

RESULT 55
AC095979
LOCUS       AC095979          235419 bp    DNA      linear    HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-11O20, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC095979
VERSION     AC095979.7  GI:24818145
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

REFERENCE   1 (bases 1 to 235419)
AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
             Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
             Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
             Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
             Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
             Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
             Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chu,J.,
             Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
             Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
             Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
             Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
             Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
             Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
             Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
             Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
             Gebregeorgis,E., Geer,K., Gill,R., Grady,W., Guerra,W., Guevara,W.,
             Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
             Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
             Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
             Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
             Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
             Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,Z., Kovar,C.,
             Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,J., Liu,J.,
             Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
             Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
             Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
             Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
             Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,M.,
             Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
             Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
             Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
             Nwaokelemeh,O., Okwuonu,G., Oiarrpunsagoon,A., Pal,S., Parks,K.,
             Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
             Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
             Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
             Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
             Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
             Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
             Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
             Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
             Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
             Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
             Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
             Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
             Williams,G., Willson,R., Wieczyk,R., Wooden,H., Worley,K.,
             Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
             Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
             Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
             Weinstock,G. and Gibbs,R.A.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 235419)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 235419)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Nov 9, 2002 this sequence version replaced gi:22772470.
             The sequence in this assembly is a combination of BAC based reads
             and whole genome shotgun sequencing reads assembled using Atlas
             (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
             in the feature table below represents a scaffold in the Atlas
             assembly (a 'contig-scaffold'). Within each contig-scaffold,
             individual sequence contigs are ordered and oriented, and separated
             by sized gaps filled with Ns to the estimated size. The sequence
             may extend beyond the ends of the clone and there may be sequence
             contigs within a contig-scaffold that consist entirely of whole
             genome shotgun sequence reads. Both end sequences and whole genome
             shotgun sequence only contigs will be indicated in the feature
             table.
             --------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
             --------- Project Information
                Center project name: GEBU
                Center clone name: CH230-11O20
             --------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 225536 bases at least Q40
                Consensus quality: 227458 bases at least Q30
                Consensus quality: 228649 bases at least Q20
                Estimated insert size: 235033; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
             ----
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             *   consists of 3 contigs. The true order of the pieces
             *   is not known and their order in this sequence record is
             *   arbitrary. Gaps between the contigs are represented as
             *   runs of N, but the exact sizes of the gaps are unknown.
             *   This record will be updated with the finished sequence
             *   as soon as it is available and the accession number will
             *   be preserved.
             *   1   221492: contig of 221492 bp in length
             *   221493 .221592: gap of unknown length
             *   221593 228651: contig of 7059 bp in length
             *   228652 228751: gap of unknown length
             *   228752 235419: contig of 6668 bp in length.
FEATURES   Location/Qualifiers
     source   1..235419
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-11O20"
     misc_feature   2958..3664
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH344795"
     misc_feature   219911..220573
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH344813"
     gap   221493..221592
                     /estimated_length=unknown
     gap   228652..228751
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:   1.88e+03   Length: 235419
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 14   Gaps: 0

US-09-989-890-238 (1-212) x AC095979 (1-235419)

Qy   60 AlaAlaGlyIleAlaSerAlaValGlu 69

Db 83366 GCTGCAGGGATTGCCTCCAGCGCTGTGGAG 83395

```
RESULT 56
SC0939120/c
LOCUS       SC0939120           248550 bp    DNA      linear   BCT 16-APR-2005
DEFINITION  Streptomyces coelicolor A3(2) complete genome; segment 17/29.
ACCESSION   AL939120 AL078618 AL160312 AL160431 AL353815 AL356334 AL392146
            AL392148 AL392175 AL392176 AL392178 AL645882
VERSION     AL939120.1  GI:24413871
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
ORGANISM    Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
AUTHORS     Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.B., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
TITLE       Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
JOURNAL     Nature 417 (6885), 141-147 (2002)
PUBMED      12000953
REFERENCE   2  (bases 1 to 248550)
AUTHORS     Bentley,S.D.
TITLE       Direct Submission
JOURNAL     Submitted (09-MAY-2002) submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 26, 2002 this sequence version replaced
            gi:11595527, gi:7242710, gi:7248323, gi:20520808, gi:7981386,
            gi:10129696, gi:10129727, gi:10178320, gi:10178340, gi:10178373.
FEATURES             Location/Qualifiers
     source          1..248550
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     gene            406..897
                     /gene="SCO4412"
                     /note="synonym: SC6F11.10"
     CDS             406..897
                     /gene="SCO4412"
                     /note="SC6F11.10, possible regulatory protein, len: 163
                     aa; similar to TR:Q9ZBT2 (EMBL:AL034446) Streptomyces
                     coelicolor putative regulatory protein SCIA9.20, 135 aa;
                     fasta scores: opt: 227 z-score: 286.6 E(): 1.8e-08; 43.0%
                     identity in 135 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein"
                     /protein_id="CAC08421.1"
                     /db_xref="GI:10178383"
                     /db_xref="GOA:Q9F376"
                     /db_xref="InterPro:IPR003594"
                     /db_xref="UniProt/TrEMBL:Q9F376"
                     /translation="MQVLQVQLEVRPDPSEVGRARRWARSRLAGSGIRADEPLAETLV
                     LLVSBLVTNAVVHTGRSAVLRLVLPGAVTEETQEATVRLEVDDASDRAPVPRCAGGES
                     TGGRGLALVDGLADRWGWSREGAGKRIWCELDRRAAGBEAPPVCGSGGGAVAYBGFAY
                     BVV"
     gene            complement(1200..2234)
                     /gene="SCO4413"
                     /note="synonym: SC6F11.11c"
     CDS             complement(1200..2234)
                     /gene="SCO4413"
                     /note="SC6F11.11c, probable araC family transcriptional
                     regulatory protein, len: 344 aa; highly similar to
                     TR:Q9SL66 (EMBL:AB023785) Streptomyces griseus AdpA, 405
                     aa; fasta scores: opt: 881 z-score: 1024.1 E(): 0; 46.9%
                     identity in 322 aa overlap and to TR:CAB70646
```

```
                     (EMBL:AL137242) Streptomyces coelicolor putative
                     transcriptional regulator SC8F4.20c, 337 aa; fasta scores:
                     opt: 1237 z-score: 1238.7 E(): 0; 60.3% identity in 325 aa
                     overlap. Contains Pfam match to entry PF00165 HTH_AraC,
                     Bacterial regulatory helix-turn-helix proteins, araC
                     family and a possible helix-turn-helix motif at residues
                     249..270 (+2.55 SD)"
                     /codon_start=1
                     /transl_table=11
                     /product="putative araC family transcriptional regulatory
                     protein"
                     /protein_id="CAC08422.1"
                     /db_xref="GI:10178384"
                     /db_xref="GOA:Q9F375"
                     /db_xref="InterPro:IPR000005"
                     /db_xref="InterPro:IPR002818"
                     /db_xref="InterPro:IPR010916"
                     /db_xref="UniProt/TrEMBL:Q9F375"
                     /translation="MAPETLSAPGAHGAPVRPHRVVVLALDGLLPFELGIASRIFGLA
                     RDEHRRPLYEVVTCSVRPPGPVATDADFPEILVGNGPEALATADTVVVPASYELGPVFE
                     EGVLTGELSAALAHVRPGTRLVAICTGGVHAAGVAAGYLDGRPATTHWASAEHFQRVFPK
                     VRVDADVLFVDDGDVLTSAGVAAGIDLCVHIVRRDHGTAVANEVARRTVVPPHRDGGQ
                     AQYIERPVPDPQQSSTTAARVWALGRLHBPIQLRDMAAQESMSVRTFTRRFRESVGVS
                     PGQWLTRQVRERARHLLESTALSVDQVAQDAGFGTAQSMRQHLQAVLGVTPGYRRTF
                     RAELSTACGB"
     misc_feature    complement(1233..1493)
                     /gene="SCO4413"
                     /note="Pfam match to entry PF00165 HTH_AraC, Bacterial
                     regulatory helix-turn-helix proteins, araC family, score
                     74.90, E-value 1.6e-18"
     RBS             2305..2308
     gene            2319..3644
                     /gene="SCO4414"
                     /note="synonym: SC6F11.12"
     CDS             2319..3644
                     /gene="SCO4414"
                     /note="SC6F11.12, possible integral membrane protein, len:
                     441 aa; similar to TR:BAB06413 (EMBL:AP001516) Bacillus
                     halodurans transporter BH2694, 418 aa; fasta scores: opt:
                     795 z-score: 909.8 E(): 0; 31.0% identity in 413 aa
                     overlap. Contains possible hydrophobic membrane spanning
                     regions"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane transport protein"
                     /protein_id="CAC08423.1"
                     /db_xref="GI:10178385"
                     /db_xref="GOA:Q9F374"
                     /db_xref="InterPro:IPR007114"
                     /db_xref="InterPro:IPR011701"
                     /db_xref="UniProt/TrEMBL:Q9F374"
                     /translation="MTQTTDAAAAAASEARPEPTRPGRPTRAHRAWVAAVFVTII
                     GAAAFRSLPGILIDPLHQEFGWNSRGTIGAAVSVNLALYGLTAPFAAALMDRFGIRRVV
                     AVALMVIALGSGLTVMNSAWQLILCWGLLVGLGGSMALAFAATVTNRWFTERRGLV
                     TGVLTAASASGOLIFLPLLSWIVETYDWRPAAVTVALALAVVPFVWLLLHDHPADVG
                     QRPYGAKEFVPRKPAPVPGAARRALTVLPAAVRTGPFWLLAGTFAICGASTNGLIQTHF
                     VPAAHDHGMPITAAASLLAVIGVFDVAGTIASGWFTDRFEPRRLLAVYYALRGVSLLF
                     LPMLLAPSVHPMIFFIVFYGLDWVATVPPTLALCREHYGBDSAIVFGWVLASHQVGA
                     ALVFALPGVARDRFGSYDVVWIASGALCAAAALMALVIRRRPVPVAALS"
     gene            3666..4538
                     /gene="SCO4415"
                     /note="synonym: SC6F11.13"
     CDS             3666..4538
                     /gene="SCO4415"
                     /note="SC6F11.13, possible integral membrane protein, len:
                     290 aa. Contains possible hydrophobic membrane spanning
                     regions"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAC08424.1"
                     /db_xref="GI:10178386"
                     /db_xref="UniProt/TrEMBL:Q9F373"
```

```
LOCUS           AC118963              249406 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION      Rattus norvegicus clone CH230-180E4, WORKING DRAFT SEQUENCE.
ACCESSION       AC118963
VERSION         AC118963.4  GI:25009106
KEYWORDS        HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE          Rattus norvegicus (Norway rat)
  ORGANISM      Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE       1  (bases 1 to 249406)
  AUTHORS       Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
                Allen,C., Allen,H., Alsbrooks,J., Amin,A., Anguiano,D.,
                Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
                Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
                Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
                Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
                Cardenas,V., Carter,K., Cavazos,I., Ceasar,A., Center,A.,
                Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J.,
                Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
                Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
                Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
                Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
                Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
                Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
                Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
                Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
                Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
                Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
                Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
                Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
                Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
                Karpathy,S., Kelly,S., Khan,J., King,L., Kovar,C.,
                Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
                Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
                Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
                Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
                Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
                Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
                Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
                Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
                Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
                Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pai,S., Pfannkoch,C.,
                Pasternak,S., Paul,H., Perez,A., Perez,L., Primus,E., Pu,L.,
                Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L-L.,
                Puazo,J., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
                Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
                Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
                Sanders,R., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
                Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
                Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
                Steimle,V., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
                Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
                Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
                Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
                Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
                Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
                Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
                Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
                Weinstock,G. and Gibbs,R.A.
  TITLE         Direct Submission
  JOURNAL       Unpublished
REFERENCE       2  (bases 1 to 249406)
  AUTHORS       Worley,K.C.
  TITLE         Direct Submission
  JOURNAL       Submitted (22-APR-2002) Human Genome Sequencing Center, Department
                of Molecular and Human Genetics, Baylor College of Medicine, One
                Baylor Plaza, Houston, TX 77030, USA
REFERENCE       3  (bases 1 to 249406)
  AUTHORS       Rat Genome Sequencing Consortium.
  TITLE         Direct Submission
  JOURNAL       Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
                of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
     gene            /translation="MGWGTEPEAAEGRPLPWSNDEWNVAAALCATQLPLMWLAWWFFV
                     GAGQDDHGRGGGAFGVLCLPLVLPLLGVHSSAQIMPAATLARLLPRRSGPEWARHHV
                     ASVLVGAGWAVLGHLVWGWPPRDTVPWFAGAGILPVLFLAPLRRRAWGGWSVWLRSAG
                     ACFVLSVAGGLAVAALADEYRPPALSDGQLAGEWRGDHGAVLRLAPGGRAGLTRVPSR
                     PDFGTYRDYTRCSGTGTWSRNLSGDREGVVVRLDGDCGEETSWTIGGSERSPELFALF
                     GDPDAGDLLFLTRN"
                     complement (4721..5926)
                     /gene="SCO4416"
                     /note="synonym: SC6F11.14c"
     CDS             complement (4721..5926)
                     /gene="SCO4416"
                     /note="SC6F11.14c, probable monooxygenase, len: 401 aa;
                     similar to TR:Q9SDE6 (EMBL:AP000837) Oryza sativa similar
                     to human dimethylanilyne monooxygenase, 437 aa; fasta
                     scores: opt: 692 z-score: 722.8 E(): 9.3e-33; 38.2%
                     identity in 403 aa overlap. Contains Pfam match to entry
                     PF00743 FMO-like, Flavin-binding monooxygenase-like"
                     /codon_start=1
                     /transl_table=11
                     /product="putative monooxygenase"
                     /protein_id="CAC08425.1"
                     /db_xref="GI:10178387"
                     /db_xref="GOA:Q9F372"
                     /db_xref="InterPro:IPR000103"
                     /db_xref="InterPro:IPR000205"
                     /db_xref="InterPro:IPR000759"
                     /db_xref="InterPro:IPR001327"
                     /db_xref="UniProt/TrEMBL:Q9F372"
                     /translation="MADSTASAEPQRPTHPAGRPVYVIGAGPGGLAVAHALRARGLRA
                     VVLERADHVGSSWRRHYDRLRLHTTRRLSALPGLPIPRRFGRWVARDDVVRYLEKYAE
                     YHQLEIVTGVEVFRVRERAPDGTGWLLHAAGRELTGAAVVATGYNHTPRVPDWPGRD
                     TYTGEFRHAADYRTPAPYAGRDVLVVGVGNTGAEIAVDLVEGGAARVRLAVRTAPHIV
                     RRSTAGWAAQYTGVLCRRLPVALVDRLARPLARISVPDLSAQGLPRPGTGLYSRVAEG
                     AIPVQDVGLIDAVRSGRVEVVAAMDGPEDGKVLLADGTRIIAPDAVIAATGYRRGLEGL
                     VGHLGVLDGTGRPVVQGGRTPAAAPGLYFTGFTNPISGMLRELAIDAERIIAGAVAKRG
                     AGGVSRLPG"
     misc_feature    complement (4754..5875)
                     /gene="SCO4416"
                     /note="Pfam match to entry PF00743 FMO-like,
                     Flavin-binding monooxygenase-like, score -255.00, E-value
                     8.8e-09"
     gene            6026..6643
                     /gene="SCO4417"
                     /note="synonym: SC6F11.15"
     CDS             6026..6643
                     /gene="SCO4417"
                     /note="SC6F11.15, probable acetyltransferase, len: 205 aa;
                     similar to SW:RIML_ECOLI (EMBL:X15860) Escherichia coli
                     ribosomal-protein-serine acetyltransferase (EC 2.3.1.-)
                     RimL, 179 aa; fasta scores: opt: 199 z-score: 261.1 E():
                     4.8e-07; 30.6% identity in 157 aa overlap. Contains Pfam
                     match to entry PF00583 Acetyltransf, Acetyltransferase
                     (GNAT) family"
                     /codon_start=1
                     /transl_table=11
                     /product="putative acetyltransferase"
```

```
Alignment Scores:
Pred. No.:                1.97e+03          Length:              248550
Score:                    10.00             Matches:             10
Percent Similarity:       100.0%            Conservative:        0
Best Local Similarity:    100.0%            Mismatches:          0
Query Match:              4.7%              Indels:              0
DB:                       1                 Gaps:                0

us-09-989-890-238 (1-212) x SCO939120 (1-248550)

Qy     51 GlyAlaGlyLeuProSerAlaSerAlaAla 60
          |||||||||||||||||||||||||||||||
Db 243017 GGGCGCCGGCTTGCCCTCGGCCTCGGCGGCC 242988

RESULT 57
, AC118963/c
```

COMMENT

Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23269749.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contig will be indicated in the feature
table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GVHY
Center clone name: CH230-180B4
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 213357 bases at least Q40
Consensus quality: 215358 bases at least Q30
Consensus quality: 217213 bases at least Q20
Estimated insert size: 223109; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*   1 249406: contig of 249406 bp in length.

FEATURES   Location/Qualifiers
source   1..249406
                /organism="Rattus norvegicus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10116"
                /clone="CH230-180B4"
misc_feature   1..1279
                /note="wgs_contig"

ORIGIN

Alignment Scores:
Pred. No.:   1.97e+03   Length: 249406
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 14   Gaps: 0

US-09-989-890-238 (1-212) x AC118963 (1-249406)

Qy   60 AlaAlaGlyIleAlaSerSerAlaValGlu 69
           ||||||||||||||||||||||||||||||
Db   14476 GCTGCAGGGAGGATTGGCCTCCAGCGCTGTGGAG 14447

RESULT 58
AC137427/c
LOCUS   AC137427   254687 bp   DNA   linear   HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-unknown, *** SEQUENCING IN PROGRESS
            ***, 7 unordered pieces.
ACCESSION   AC137427
VERSION   AC137427.1   GI:25138551
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 254687)
AUTHORS   Muzny,D.Marie, Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,H., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Derano,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,H., Mahmood,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,B.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
            Milosavljevic,A., Miner,G., Minja,B., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olasrpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,F., Quiroz,J., Rachlin,E., Reeves,M., Regler,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Shen,H.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wel,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 254687)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole

```
Db       243241 CCATCCAGCTTAGCTCTTCTCCAGACCC 243212
                |||||||||||||||||||||||||||||||||||||||||
RESULT 59
AC126312/c
LOCUS       AC126312            256017 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-1B9, *** SEQUENCING IN PROGRESS ***,
            3 unordered pieces.
ACCESSION   AC126312
VERSION     AC126312.4  GI:24941465
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 256017)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,J., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 256017)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 256017)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23267435.
            The sequence in this assembly is a combination of BAC based reads
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            --------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------------- Project Information
            Center project name: KZLI
            Center clone name: CH230-unknown
            --------------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 206091 bases at least Q40
            Consensus quality: 212900 bases at least Q30
            Consensus quality: 217431 bases at least Q20
            Estimated insert size: 219136; sum-of-contigs estimation
            Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1  32501; contig of 32501 bp in length
            *    32502  .32601
            *    32601  32601; gap of unknown length
            *    32602  44171; contig of 11570 bp in length
            *    44171  44271; gap of unknown length
            *    44172  .44271
            *    44206  245205; contig of 200934 bp in length
            *   245205  .245305; gap of unknown length
            *   245206  .245305
            *   245306  246951; contig of 1646 bp in length
            *   246952  247051
            *   247051  247051; gap of unknown length
            *   247052  248359; contig of 1308 bp in length
            *   248359  248459; gap of unknown length
            *   248360  248459
            *   248459  251483; contig of 3024 bp in length
            *   251483  251583; gap of unknown length
            *   251484  .251583
            *   251584  254687; contig of 3104 bp in length.
FEATURES             Location/Qualifiers
     source          1..254687
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-unknown"
     gap             32502..32601
                     /estimated_length=unknown
     gap             44172..44271
                     /estimated_length=unknown
     gap             245206..245305
                     /estimated_length=unknown
     gap             246952..247051
                     /estimated_length=unknown
     gap             248360..248459
                     /estimated_length=unknown
     gap             251484..251583
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:               2.0e+03       Length:           254687
Score:                   10.00         Matches:          10
Percent Similarity:      100.0%        Conservative:     0
Best Local Similarity:   100.0%        Mismatches:       0
Query Match:             4.7%          Indels:           0
DB:                      14            Gaps:             0

US-09-989-890-238 (1-212) x AC137427 (1-254687)

Qy       25 ProSerSerLeuAlaLeuSerArgPro 34
```

```
Best Local Similarity:  100.0%        Mismatches:    0
Query Match:             4.7%         Indels:        0
DB:                      14           Gaps:          0

US-09-989-890-238 (1-212) x AC126312 (1-256017)

Qy       60 AlaAlaGlyIleAlaSerSerAlaValGlu 69
            ||||||||||||||||||||||||||||||
Db   155464 GCTGCAGGGATTGCCTCCAGCGCTGGGAG 155435

RESULT 60
AC107549
LOCUS       AC107549      346316 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-45E20, *** SEQUENCING IN PROGRESS
            ***,11 unordered pieces.
ACCESSION   AC107549
VERSION     AC107549.4  GI:2326118
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 346316)
AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,M., Ayodeji,M., Baca,B., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,M., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton.N., Nguyen.N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,M., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,B., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,H., Wleczyk,R., Wooden,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
```

The DNA sequence reported in this record was derived by a
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

```
            ---------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
            Center project name: TUNJ
            Center clone name: CH230-1B9
            ---------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 214827 bases at least Q40
            Consensus quality: 219116 bases at least Q30
            Consensus quality: 221911 bases at least Q20
            Estimated insert size: 222841; sum-of-contigs estimation
            Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

          *     1     83721: contig of 83721 bp in length
          * 83722     83821: gap of unknown length
          * 83822   253567: contig of 16974G bp in length
          * 253567   253667: gap of unknown length
          * 253568   256017: contig of 2350 bp in length.

          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          *   consists of 3 contigs. The true order of the pieces
          *   is not known and their order in this sequence record is
          *   arbitrary. Gaps between the contigs are represented as
          *   runs of N, but the exact sizes of the gaps are unknown.
          *   This record will be updated with the finished sequence
          *   as soon as it is available and the accession number will
          *   be preserved.
FEATURES             Location/Qualifiers
     source          1..256017
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-1B9"
     misc_feature    6668..7475
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH274997"
     misc_feature    38358..39527
                     /note="wgs_contig"
     misc_feature    67950..69005
                     /note="wgs_contig"
     misc_feature    78781..81213
                     /note="wgs_contig"
     gap             83722..83821
                     /estimated_length=unknown
     misc_feature    83822..85574
                     /note="wgs_contig"
     misc_feature    219978..221353
                     /note="wgs_contig"
     gap             253568..253667
                     /estimated_length=unknown
ORIGIN
```

```
Alignment Scores:
Pred. No.:               2.02e+03        Length:          256017
Score:                   10.00           Matches:         10
Percent Similarity:      100.0%          Conservative:    0
```

```
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 346316)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3  (bases 1 to 346316)
AUTHORS      Rat Genome Sequencing Consortium.
TITLE        Direct Submission
JOURNAL      Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Sep 22, 2002 this sequence version replaced gi:21731862.
             The sequence in this assembly is a combination of BAC based reads
             and whole genome shotgun sequencing reads assembled using Atlas
             (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
             sequence may extend beyond the ends of the clone and there may be
             contigs that consist entirely of whole genome shotgun sequence
             reads.  Both end sequences and whole genome shotgun sequence only
             contig will be indicated in the feature table.
             -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
             -------------- Project information
               Center project name: GLDB
               Center clone name: CH230-45E20
             -------------- Summary Statistics
               Assembly program: Phrap; version 0.990329
               Consensus quality: 188416 bases at least Q40
               Consensus quality: 193644 bases at least Q30
               Consensus quality: 196389 bases at least Q20
               Estimated insert size: 218935; sum-of-contigs estimation
               Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
             ---------------
           * NOTE: Estimated insert size may differ from sequence length.
           *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html).
           * NOTE: This is a 'working draft' sequence. It currently
           *   consists of 11 contigs. The true order of the pieces
           *   is not known and their order in this sequence record is
           *   arbitrary. Gaps between the contigs are represented as
           *   runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           *   as soon as it is available and the accession number will
           *   be preserved.
           *         1          225315: contig of 225315 bp in length
           *    225316          225415: gap of unknown length
           *    225416          246434: contig of 21019 bp in length
           *    246435          246534: gap of unknown length
           *    246535          310670: contig of 64136 bp in length
           *    310671          310770: gap of unknown length
           *    310771          312038: contig of 1268 bp in length
           *    312039          312138: gap of unknown length
           *    312139          314414: contig of 2276 bp in length
           *    314415          314514: gap of unknown length
           *    314515          315550: contig of 1036 bp in length
           *    315551          315650: gap of unknown length
           *    315651          317317: contig of 1667 bp in length
           *    317318          317417: gap of unknown length
           *    317418          328186: contig of 10769 bp in length
           *    328187          328286: gap of unknown length
           *    328287          332953: contig of 4667 bp in length
           *    332954          333053: gap of unknown length
           *    333054          342104: contig of 9051 bp in length
           *    342105          342204: gap of unknown length
           *    342205          346316: contig of 4112 bp in length.
FEATURES             Location/Qualifiers
     source          1..346316
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:10116"
                     /clone="CH230-45E20"
     misc_feature    1..1513
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    393..991
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH297453"
     misc_feature    2317..2759
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH297451"
     gap             225316..225415
                     /estimated_length=unknown
     gap             246435..246534
                     /estimated_length=unknown
     gap             310671..310770
                     /estimated_length=unknown
     gap             312039..312138
                     /estimated_length=unknown
     gap             314415..314514
                     /estimated_length=unknown
     gap             315551..315650
                     /estimated_length=unknown
     gap             317318..317417
                     /estimated_length=unknown
     gap             328187..328286
                     /estimated_length=unknown
     gap             332954..333053
                     /estimated_length=unknown
     gap             342105..342204
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                                Length:          346316
Score:             2.61e+03               Matches:         10
Percent Similarity: 10.00%                Conservative:    0
Best Local Similarity: 100.0%             Mismatches:      0
Query Match:       4.7%                   Indels:          0
DB:                14                     Gaps:            0

US-09-989-890-238 (1-212) x AC107549 (1-346316)

Qy          25 ProSerSerLeuAlaLeuLeuSerArgPro 34
                  |||||||||||||||||||||||||||||
Db      278434 CCATCCAGCTTAGCTCTTCTCCAGACCC 278463

RESULT 61
AB032303/c
LOCUS       AB032303                100 bp    DNA        linear   VRL 25-JAN-2000
DEFINITION  TT virus DNA, untranslated region, isolate:Mf-TTV5.
ACCESSION   AB032303
VERSION     AB032303.1  GI:6751867
KEYWORDS    .
SOURCE      Torque teno virus (TTV)
  ORGANISM  Torque teno virus
            Viruses; ssDNA viruses; Anelloviridae.
REFERENCE   1  (sites)
  AUTHORS   Okamoto,H., Fukuda,M., Tawara,A., Nishizawa,T., Itoh,Y.,
            Hayasaka,I., Tsuda,F., Tanaka,T., Miyakawa,Y. and Mayumi,M.
  TITLE     Species-specific TT viruses and cross-species infection in nonhuman
            primates
  JOURNAL   J. Virol. 74 (3), 1132-1139 (2000)
   PUBMED   10627523
REFERENCE   2  (bases 1 to 100)
  AUTHORS   Okamoto,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1999) Hiroaki Okamoto, Jichi Medical School,
            Immunology Division; 3311-1 Yakushiji, Minamikawachi-machi,
```

```
                  Kawachi-gun, Tochigi 329-0498, Japan (E-mail:hokamoto@jichi.ac.jp,
                  Tel:81-285-58-7404, Fax:81-285-44-1557)
FEATURES             Location/Qualifiers
     source          1..100
                     /organism="Torque teno virus"
                     /mol_type="genomic DNA"
                     /isolate="Mf-TTV5"
                     /specific_host="Macaca fuscata"
                     /db_xref="taxon:68887"
     misc_feature    1..100
                     /note="untranslated region"
ORIGIN

Alignment Scores:
Pred. No.:
Score:                  21.7        Length:           100
                         9.00       Matches:          9
Percent Similarity:     100.0%      Conservative:     0
Best Local Similarity:  100.0%      Mismatches:       0
Query Match:            4.2%        Indels:           0
DB:                     13          Gaps:             0

US-09-989-890-238 (1-212) x AB032303 (1-100)

Qy   26 SerSerLeuAlaLeuSerArgPro 34
        |||||||||||||||||||
Db   45 TCGTCACTGGCGCTCCTCTCCCGACCT 19


RESULT 62
AB032304/c
LOCUS       AB032304              100 bp    DNA     linear   VRL 25-JAN-2000
DEFINITION  TT virus DNA, untranslated region, isolate:Mf-TTV6.
ACCESSION   AB032304
VERSION     AB032304.1  GI:6751868
KEYWORDS
SOURCE      Torque teno virus (TTV)
  ORGANISM  Torque teno virus
            Viruses; ssDNA viruses; Anellovirus.
REFERENCE   1 (sites)
  AUTHORS   Okamoto,H., Fukuda,M., Tawara,A., Nishizawa,T., Itoh,Y.,
            Hayasaka,I., Tsuda,F., Tanaka,T., Miyakawa,Y. and Mayumi,M.
  TITLE     Species-specific TT viruses and cross-species infection in nonhuman
            primates
  JOURNAL   J. Virol. 74 (3), 1132-1139 (2000)
   PUBMED   10627523
REFERENCE   2 (bases 1 to 100)
  AUTHORS   Okamoto,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1999) Hiroaki Okamoto, Jichi Medical School,
            Immunology Division; 3311-1 Yakushiji, Minamikawachi-machi,
            Kawachi-gun, Tochigi 329-0498, Japan (E-mail:hokamoto@jichi.ac.jp,
            Tel:81-285-58-7404, Fax:81-285-44-1557)
FEATURES             Location/Qualifiers
     source          1..100
                     /organism="Torque teno virus"
                     /mol_type="genomic DNA"
                     /isolate="Mf-TTV6"
                     /specific_host="Macaca fuscata"
                     /db_xref="taxon:68887"
     misc_feature    1..100
                     /note="untranslated region"
ORIGIN

Alignment Scores:
Pred. No.:
Score:                  21.7        Length:           100
                         9.00       Matches:          9
Percent Similarity:     100.0%      Conservative:     0
Best Local Similarity:  100.0%      Mismatches:       0
Query Match:            4.2%        Indels:           0
DB:                     13          Gaps:             0

US-09-989-890-238 (1-212) x AB032304 (1-100)

Qy   26 SerSerLeuAlaLeuSerArgPro 34
        |||||||||||||||||||
Db   45 TCGTCACTGGCGCTCCTCTCCCGACCT 19


RESULT 63
HS295374/c
LOCUS       HS295374              140 bp    DNA     linear   PRI 02-DEC-1997
DEFINITION  H.sapiens genomic DNA fragment (clone NRLA100R).
ACCESSION   295374
VERSION     295374.1  GI:2661928
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Gizatullin,R.Z., Kashuba,V.I., Protopopov,A.I. and Zabarovsky,E.R.
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 140)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-1997) Zabarovsky E.R., Karolinska Institute,
            Microbiology and Tumorbiology Center, P.O. Box 280, Stockholm,
            S-171 77, Sweden
FEATURES             Location/Qualifiers
     source          1..140
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="NRLA100R"
                     /clone_lib="NRL NotI linking library"
                     /lab_host="mouse/human microcell hybrid MCH903.1"
     misc_feature    1..140
                     /note="sequence surrounding NotI site"
ORIGIN

Alignment Scores:
Pred. No.:
Score:                  29          Length:           140
                         9.00       Matches:          9
Percent Similarity:     100.0%      Conservative:     0
Best Local Similarity:  100.0%      Mismatches:       0
Query Match:            4.2%        Indels:           0
DB:                     8           Gaps:             0

US-09-989-890-238 (1-212) x HS295374 (1-140)

Qy   37 ProProProAlaCysSerGlyAsp 45
        ||||||||||||||||||||||
Db   83 CCGCCGCCCCGCCGCGTGTTCCGGAGAC 57


RESULT 64
AY745220
LOCUS       AY745220              303 bp    mRNA    linear   INV 09-OCT-2004
DEFINITION  Anopheles gambiae cytochrome P450 (CYP49A1) mRNA, partial cds.
ACCESSION   AY745220
VERSION     AY745220.1  GI:53766654
KEYWORDS
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Anophelinae; Anopheles.
REFERENCE   1 (bases 1 to 303)
  AUTHORS   Ranson,H. and Nikou,D.
  TITLE     Cytochrome P450s from the malaria vector, Anopheles gambiae
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 303)
  AUTHORS   Ranson,H. and Nikou,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-2004) Vector Research, Liverpool School of
            Tropical Medicine, Pembroke Place, Liverpool, Merseyside L3 5QA, UK
FEATURES             Location/Qualifiers
```

source   1..303
                /organism="Anopheles gambiae"
                /mol_type="mRNA"
                /db_xref="taxon:7165"
                /dev_stage="larvae"
gene   <1..>303
                /gene="CYP49A1"
CDS   <1..>303
                /gene="CYP49A1"
                /codon_start=1
                /product="cytochrome P450"
                /protein_id="AAU93487.1"
                /db_xref="GI:53766655"
                /translation="FPHLVVSNLQQYFPEPDRFVPERWLKRGELKEHSGCPHAGQKIH
                PYVSLPFGYGRRTCIGRRFABCELQILLSKLFRRYQVEYNYBKLTYKVNPTYIPDKP"

ORIGIN

Alignment Scores:
Pred. No.:   56.2   Length: 303
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 2   Gaps: 0

US-09-989-890-238 (1-212) x AY745220 (1-303)

Qy   50 SerGlyAlaGlyLeuProSerAlaSer 58
           |||||||||||||||||||
Db   176 TCGGGCGCCGGTTTGCCCAGTGCGAGC 202

RESULT 65
BV106213/c
LOCUS   BV106213   340 bp   DNA   linear   STS 05-MAR-2004
DEFINITION MARC 3519-3520:996679156:1 RTS-1 Bos indicus x Bos taurus STS
            genomic, sequence tagged site.
ACCESSION   BV106213
VERSION   BV106213.1   GI:45184162
KEYWORDS   STS.
SOURCE   Bos indicus x Bos taurus (hybrid cattle)
ORGANISM   Bos indicus x Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 340)
AUTHORS   Snelling,W.M., Casas,E., Stone,R.T., Smith,T.P.L., Keele,J.W. and
            Bennett,G.L.
TITLE   Linkage mapping bovine ESTs
JOURNAL   Unpublished (2004)
COMMENT

            Contact: Stone RT
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4166
            Fax: 402 762 4173
            Email: stone@email.marc.usda.gov
            Primer A: AGATGTTTCTTAACACGCTGGA
            Primer B: GGAACCACCAAAGTTACAAGC
            STS size: 500
            PCR Profile:
                Hotstart: 95 degrees for 15 minutes
                Denature: 95 degrees for 30 seconds
                Anneal: 58 degrees for 30 seconds
                Extension: 68 degrees for 1 minute
                Cycles: 32 to 45
            Protocol:
                Template: 25-50 ng genomic DNA
                Primer: each 0.4 uM
                dNTPs: each 100 uM
                Taq Polymerase: 0.25 units (Qiagen HotStar)

Buffer: Commercially supplied Qiagen HotStar buffer

            The STS is derived from PCR amplicons generated from genomic DNA,
            sequenced from each end using the amplification primers. The
            sequence does not necessarily represent the entire amplicon.
            Sequence derived from PolyPhred was trimmed from each end of each
            unique contig until five consecutive bases exceeded a quality score
            threshold of 20, and the next ten bases averaged a quality score of
            20 or greater. Amplicon size was estimated by agarose gel
            electrophoresis.

FEATURES   Location/Qualifiers
source   1..340
                /organism="Bos indicus x Bos taurus"
                /mol_type="genomic DNA"
                /db_xref="taxon:30522"
                /clone_lib="RTS-1"
                /dev_stage="adult"
                /note="taurus and indicus crossbreds"
STS   <1..>340
ORIGIN

Alignment Scores:
Pred. No.:   62.1   Length: 340
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 10   Gaps: 0

US-09-989-890-238 (1-212) x BV106213 (1-340)

Qy   47 GlyCysGlySerGlyAlaGlyLeuPro 55
           |||||||||||||||||||||
Db   44 GGTTGCGGTAGTGTGCAGGTCTCCA 18

RESULT 66
G67666/c
LOCUS   G67666   354 bp   DNA   linear   STS 30-JUN-2004
DEFINITION MARC3174-3175 Bovine white blood cells Bos taurus STS genomic
            sequence tagged site.
ACCESSION   G67666
VERSION   G67666.1   GI:12802954
KEYWORDS   STS.
SOURCE   Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 354)
AUTHORS   Stone,R.T., Grosse,W.M., Casas,E., Smith,T.P., Keele,J.W. and
            Bennett,G.L.
TITLE   Use of bovine EST data and human genomic sequences to map 100
            gene-specific bovine markers
JOURNAL   Mamm. Genome 13 (4), 211-215 (2002)
PUBMED   11956765
COMMENT

            Contact: Timothy P.L. Smith
            USDA/ARS U.S. Meat Animal Research Center
            Box 166, Spur 18d, Clay Center, NE 68933, USA
            Email: smith@email.marc.usda.gov
            Primer A: GCCTGTACATCTGGACTTTCG
            Primer B: TCCAGTGAGTAGGTGACCCC
            PCR Profile:
                Hotstart: 15 minutes 95 degrees
                Denature: 30 seconds 95 degrees
                Anneal: 30 seconds 58 degrees
                Extension: 2 minutes 68 degrees
            Protocol:
                Template: 50 ng genomic DNA
                Primer: each pmoles
                dNTPs: each 100 uM
                Taq Polymerase: 0.25 units

Buffer:

Buffer:
　　Commercially supplied Qiagen HotStar buffer

The STS sequence represents a contig of the sequences derived from each end using the amplification primers to sequence. The sequence does not represent the entire amplicon, as sequence close to the primers was of low quality.

```
FEATURES             Location/Qualifiers
     source          1..354
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /clone_lib="taxon:9913"
                     /db_xref="taxon:9913"
                     /dev_stage="adult"
                     /note="PCR amplification product from genomic DNA isolated
                     from bovine WBC, Mixed breed crosses."
     STS             <1..353
     primer_bind     complement(334..353)
ORIGIN
```

```
Alignment Scores:
Pred. No.:            64.2        Length:          354
Score:                9.00        Matches:           9
Percent Similarity:  100.0%       Conservative:      0
Best Local Similarity: 100.0%     Mismatches:        0
Query Match:          4.2%        Indels:            0
DB:                   10          Gaps:              0
```

US-09-989-890-238 (1-212) x G67666 (1-354)

```
Qy   47 GlyCysGlySerGlyAlaGlyLeuPro 55
        |||||||||||||||||||||||
Db   57 GGTTGCGGTAGTGGTGCAGGTCTTCCA 31
```

```
RESULT 67
BV104173/c
LOCUS       BV104173               354 bp    DNA     linear   STS 05-MAR-2004
DEFINITION  MARC 3174-3175:96689 4323:1 RTS-1 Bos indicus x Bos taurus STS
            genomic, sequence tagged site.
ACCESSION   BV104173
VERSION     BV104173.1  GI:4518 2089
KEYWORDS    STS.
SOURCE      Bos indicus x Bos taurus (hybrid cattle)
  ORGANISM  Bos indicus x Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Snelling,W.M., Casas,E., Stone,R.T., Smith,T.P.L., Keele,J.W. and
            Bennett,G.L.
  TITLE     Linkage mapping bovine ESTs
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stone RT
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4166
            Fax: 402 762 4173
            Email: stone@email.marc.usda.gov
            Primer A: GCCTGTACATCTGGACTTTCG
            Primer B: TCCAGTGAGTAGGTGACCCC
            STS size: 400
            PCR Profile:
             Hotstart:   95 degrees for 15 minutes
             Denature:   95 degrees for 30 seconds
             Anneal:     58 degrees for 30 seconds
             Extension:  68 degrees for 1 minute
             Cycles:     32 to 45
            Protocol:
             Template:   25-50 ng genomic DNA
             Primer:     each 0.4 uM
             dNTPs:      each 100 uM
             Taq Polymerase: 0.25 units (Qiagen HotStar)
```

Buffer:
　　Commercially supplied Qiagen HotStar buffer

The STS is derived from PCR amplicons generated from genomic DNA, sequenced from each end and using the amplification primers. The sequence does not necessarily represent the entire amplicon. Sequence derived from PolyPhred was trimmed from each end of each unique contig until five consecutive bases exceeded a quality score of threshold of 20, and the next ten bases averaged a quality score of 20 or greater. Amplicon size was estimated by agarose gel electrophoresis.

```
FEATURES             Location/Qualifiers
     source          1..354
                     /organism="Bos indicus x Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:30522"
                     /clone_lib="RTS-1"
                     /dev_stage="adult"
                     /note="taurus and indicus crossbreeds"
     STS             <1..353
ORIGIN               complement(334..353)
```

```
Alignment Scores:
Pred. No.:            64.2        Length:          354
Score:                9.00        Matches:           9
Percent Similarity:  100.0%       Conservative:      0
Best Local Similarity: 100.0%     Mismatches:        0
Query Match:          4.2%        Indels:            0
DB:                   10          Gaps:              0
```

US-09-989-890-238 (1-212) x BV104173 (1-354)

```
Qy   47 GlyCysGlySerGlyAlaGlyLeuPro 55
Db   57 GGTTGCGGTAGTGGTGCAGGTCTTCCA 31
```

```
RESULT 68
BV104270/c
LOCUS       BV104270               379 bp    DNA     linear   STS 05-MAR-2004
DEFINITION  MARC 3198-3199:96688 1434:1 RTS-1 Bos indicus x Bos taurus STS
            genomic, sequence tagged site.
ACCESSION   BV104270
VERSION     BV104270.1  GI:4518 2188
KEYWORDS    STS.
SOURCE      Bos indicus x Bos taurus (hybrid cattle)
  ORGANISM  Bos indicus x Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Snelling,W.M., Casas,B., Stone,R.T., Smith,T.P.L., Keele,J.W. and
            Bennett,G.L.
  TITLE     Linkage mapping bovine ESTs
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stone RT
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4166
            Fax: 402 762 4173
            Email: stone@email.marc.usda.gov
            Primer A: ACAAGAGAACGATGCAGGG
            Primer B: GGGAGATCACCTCTCGATCA
            STS size: 500
            PCR Profile:
             Hotstart:   95 degrees for 15 minutes
             Denature:   95 degrees for 30 seconds
             Anneal:     58 degrees for 30 seconds
             Extension:  68 degrees for 1 minute
             Cycles:     32 to 45
            Protocol:
             Template:   25-50 ng genomic DNA
```

Buffer:
    Commercially supplied Qiagen HotStar buffer

Primer: each 0.4 uM
dNTPs: each 100 uM
Taq Polymerase: 0.25 units (Qiagen HotStar)

The STS is derived from PCR amplicons generated from genomic DNA,
sequenced from each end using the amplification primers. The
sequence does not necessarily represent the entire amplicon.
Sequence derived from PolyPhred was trimmed from each end of each
unique contig until five consecutive bases exceeded a quality score
of 20 or greater, and the next ten bases averaged a quality score of
20 or greater. Amplicon size was estimated by agarose gel
electrophoresis.

FEATURES   Location/Qualifiers
source   1..379
             /organism="Bos indicus x Bos taurus"
             /mol_type="genomic DNA"
             /db_xref="taxon:30522"
             /clone_lib="RTS-1"
             /dev_stage="adult"
             /note="taurus and indicus crossbreds"
STS   <1..>379
ORIGIN

Alignment Scores:
Pred. No.:   Length: 379
Score: 68.1   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 10   Gaps: 0

US-09-989-890-238 (1-212) x BV104270 (1-379)

Qy 47 GlyCysGlySerGlyAlaGlyLeuPro 55
       |||||||||||||||||||||||
Db 82 GGTTGCGGTAGTGTGTGCGAGGTCTTCCA 56

RESULT 69
AF402268
LOCUS   AF402268   387 bp   DNA   linear   BCT 10-SEP-2002
DEFINITION Anaplasma marginale expression site variant 196R-D major surface
            protein 2 (msp2) gene, partial cds.
ACCESSION   AF402268
VERSION   AF402268.1 GI:22037252
KEYWORDS   .
SOURCE   Anaplasma marginale str. Florida
  ORGANISM Anaplasma marginale str. Florida
            Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
            Anaplasmataceae; Anaplasma.
REFERENCE   1 (bases 1 to 387)
  AUTHORS   Brayton,K.A., Palmer,G.H., Lundgren,A., Yi,J. and Barbet,A.F.
  TITLE   Antigenic variation of Anaplasma marginale msp2 occurs by
            combinatorial gene conversion
  JOURNAL   Mol. Microbiol. 43 (5), 1151-1159 (2002)
  PUBMED   11918803
REFERENCE   2 (bases 1 to 387)
  AUTHORS   Brayton,K.A., Palmer,G.H., Lundgren,A., Yi,J. and Barbet,A.F.
  TITLE   Direct Submission
  JOURNAL   Submitted (25-JUL-2001) Veterinary Microbiology and Pathology,
            Washington State University, Pullman, WA 99164-7040, USA
FEATURES   Location/Qualifiers
source   1..387
            /organism="Anaplasma marginale str. Florida"
            /mol_type="genomic DNA"
            /strain="Florida"
            /isolate="expression site variant 196R-D"
            /db_xref="taxon:320483"
gene   <1..>387
            /gene="msp2"
CDS   <1..>387
            /gene="msp2"
            /note="MSP2"
            /codon_start=1
            /transl_table=11
            /product="major surface protein 2"
            /protein_id="AAM89963.1"
            /db_xref="GI:22037253"
            /translation="LGKELAYDTARGQVDRLATALGKWTKGEAKKWGNAIESATGTTS
            GDELSKKVCGKGTTSGSTNQCGVNATSGSTNNGKLSTVFNTDGABAISSMDTTASGTS
            STISLQGMAGNINSLTKDEKAIVAGAF"
ORIGIN

Alignment Scores:
Pred. No.:   Length: 387
Score: 69.3   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 1   Gaps: 0

US-09-989-890-238 (1-212) x AF402268 (1-387)

Qy 171 ProLeuProValLeuValAlaPro 179
        ||||||||||||||||||||||||
Db 287 CCACTGCCAGTGGTACTAGTAGCACCA 313

RESULT 70
DM190E7T
LOCUS   DM190E7T   442 bp   DNA   linear   STS 06-MAY-1998
DEFINITION Drosophila melanogaster STS determined from European Mapping
            Project cosmid, sequence tagged site.
ACCESSION   AL023365
VERSION   AL023365.1 GI:3123417
KEYWORDS   STS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 442)
  AUTHORS   Papagiannakis,G., Spanos,L., Cox,S., Siden-Kiamos,I. and Louis,C.
  TITLE   Sequencing the distal X chromosome of Drosophila melanogaster
  JOURNAL   Unpublished
  REMARK   Institute of Molecular Biology and Biotechnology / Fo.R.T.H., P.O.
            Box 1527, GR-711 10 Heraklion, Crete, Greece.
REFERENCE   2 (bases 1 to 442)
  AUTHORS   Ashburner,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-APR-1998) Michael Ashburner, EMBL Outstation - The
            EBI, Hinxton, Cambridge, CB10 1SD, U.K. E-mail:
            ashburner@ebi.ac.uk. On behalf of the European Drosophila Genome
            Sequencing Consortium. For further information see the European
            Drosophila Genome Sequencing Consortium's web site:
            http://edgp.ebi.ac.uk/
  REMARK   European Drosophila Genome Sequencing Consortium
COMMENT   STS_name = Dm190E7T
            clone_name = 190E7
            STS_from_promoter = T7
            vector_class = cosmid, Lorist 6
            origin_of_clone = Oregon-R
            in_situ_site_primary = 1E.
FEATURES   Location/Qualifiers
source   1..442
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /strain="Oregon-R"
            /db_xref="taxon:7227"
            /clone="190E7"
ORIGIN

Alignment Scores:
Pred. No.:   Length: 442
Score: 77.7   Matches: 9

```
Percent Similarity:     100.0%      Conservative:    0
Best Local Similarity:  100.0%      Mismatches:      0
Query DB:               4.2%        Indels:          0
DB:                     10          Gaps:            0

US-09-989-890-238 (1-212) x DM190B7T (1-442)

Qy   54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
        |||||||||||||||||||||||||||||||
Db  330 CTACCNTCCGCCAGCGCCGCCGCCGGT 356
```

RESULT 71
DM96A7T/c
```
LOCUS       DM96A7T          453 bp    DNA     linear   STS 06-MAY-1998
DEFINITION  Drosophila melanogaster STS determined from European Mapping
            Project cosmid, sequence tagged site.
ACCESSION   AL023471
VERSION     AL023471.1   GI:3123503
KEYWORDS    STS.
SOURCE      Drosophila melanogaster
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 453)
AUTHORS     Papagiannakis,G., Spanos,L., Cox,S., Siden-Kiamos,I. and Louis,C.
TITLE       Sequencing the distal X chromosome of Drosophila melanogaster
JOURNAL     Unpublished
REMARK      Institute of Molecular Biology and Biotechnology / Fo.R.T.H., P.O.
            Box 1527, GR-711 10 Heraklion, Crete, Greece.
REFERENCE   2  (bases 1 to 453)
AUTHORS     Ashburner,M.
TITLE       Direct Submission
JOURNAL     Submitted (29-APR-1998) Michael Ashburner, EMBL Outstation - The
            EBI, Hinxton, Cambridge, CB10 1SD, U.K. E-mail:
            ashburner@ebi.ac.uk. On behalf of the European Drosophila Genome
            Sequencing Consortium. For further information see the European
            Drosophila Genome Sequencing Consortium's web site:
            http://edgp.ebi.ac.uk/
REMARK      European Drosophila Genome Sequencing Consortium
COMMENT     STS_name = DM96A7T
            clone_name = 96A7
            STS_from promoter = T7
            vector_class = cosmid, Lorist 6
            origin_of_clone = Oregon-R
            in_situ_site secondary = 3C
            in_situ_repetitive_number = >40.
FEATURES             Location/Qualifiers
     source          1..453
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /strain="Oregon-R"
                     /db_xref="taxon:7227"
                     /clone="96A7"
ORIGIN
```

```
Alignment Scores:
Pred. No.:             79.4         Length:          453
Score:                 9.00        Matches:         9
Percent Similarity:    100.0%      Conservative:    0
Best Local Similarity: 100.0%      Mismatches:      0
Query DB:              4.2%        Indels:          0
DB:                    10          Gaps:            0

US-09-989-890-238 (1-212) x DM96A7T (1-453)

Qy   54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
        |||||||||||||||||||||||||||||||
Db  298 CTACCATCCGCCAGCGCCGCCGCCGGN 272
```

RESULT 72
BV308315/c
```
LOCUS       BV308315         501 bp    DNA     linear   STS 25-JAN-2005
DEFINITION  S236P6209FH1.T0 AlaskanMalamute Canis familiaris STS genomic,
            sequence tagged site.
ACCESSION   BV308315
VERSION     BV308315.1   GI:57506807
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 501)
AUTHORS     Lindblad-Toh,K.
TITLE       The genome sequence of Canis familiaris
JOURNAL     Unpublished (2004)
COMMENT     Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 501
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poodle sequence was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1..501
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
```

```
                    STS
                    ORIGIN         /strain="AlaskanMalamute"
                                   /db_xref="taxon:9615"
                                   /map="+ 23 26-445 55235003-55234584"
                                   /clone_lib="AlaskanMalamute"
                                   <1..>501

Alignment Scores:
Pred. No.:               86.5          Length:          501
Score:                   9.00          Matches:         9
Percent Similarity:      100.0%        Conservative:    0
Best Local Similarity:   100.0%        Mismatches:      0
Query Match:             4.2%          Indels:          0
DB:                                    Gaps:            0

US-09-989-890-238 (1-212) x BV308315 (1-501)

Qy    33 ArgProLeuSerProProProAlaAla 41
         |||||||||||||||||||||
Db   105 CGCCCCCTGTCGCCCCCTCCCGCGCT 79

RESULT 73
BD078483/c
LOCUS         BD078483          546 bp     DNA       linear    PAT 27-AUG-2002
DEFINITION    101 human secretory proteins.
ACCESSION     BD078483
VERSION       BD078483.1   GI:22624086
KEYWORDS      JP 2001519156-A/72.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominidae; Homo.
REFERENCE     1  (bases 1 to 546)
AUTHORS       Duan,R.D., Florence,K.A., Rosen,C.A., Ruben,S.M., Greene,J.M.,
              Young,P., Ferrie,A.M., Yu,G.L., Janat,F., Ni,J., Carter,K.C.,
              Endress,G.A., Feng,P., LaFleur,D.W. and Shi,Y.
TITLE         101 human secretory proteins
JOURNAL       Patent: JP 2001519156-A 72 23-OCT-2001;
              HUMAN GENOME SCIENCES INC
COMMENT       OS   Homo sapiens (human)
              PN   JP 2001519156-A/72
              PD   23-OCT-2001
              PF   01-OCT-1998 JP 2000515006
              PR   02-OCT-1997 US        60/060837,02-OCT-1997 US    60/060862 PR
              02-OCT-1997 US            60/060839,02-OCT-1997 US     60/060866 PR
              02-OCT-1997 US            60/060843,02-OCT-1997 US     60/060836 PR
              02-OCT-1997 US            60/060838,02-OCT-1997 US     60/060874 PR
              02-OCT-1997 US            60/060833,02-OCT-1997 US     60/060884 PR
              02-OCT-1997 US            60/060880
              PI   ROXANNE D DUAN, KIMBERLY A FLORENCE, CRAIG A ROSEN, STEVEN M  PI
              RUBEN,
              PI   JOHN M GREENE, PAUL YOUNG, ANN M FERRIE, GUO
              LIANG YU, FOUAD JANAT,
              PI   JIAN NI,
              PI   KENNETH C CARTER, GREGORY A ENDRESS, PING FENG, DAVID W LAFLEUR,
              PI   YANGGU SHI
              PC   C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P25/00,A61P35/00,
              PC   A61P37/00,
              PC   A61P43/00,C07K14/47,C07K16/00,C12N5/10,C12P21/02,C12Q1/68,  PC
              G01N33/53,
              PC   G01N33/53,G01N33/566,C12N15/00,A61K37/02,C12N5/00 CC   101
              human secretory proteins
FH   Key             Location/Qualifiers
FT   source          1..546
FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1..546
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
Alignment Scores:
Pred. No.:               93.2          Length:          546
Score:                   9.00          Matches:         9
Percent Similarity:      100.0%        Conservative:    0
Best Local Similarity:   100.0%        Mismatches:      0
Query Match:             4.2%          Indels:          0
DB:                                    Gaps:            0

US-09-989-890-238 (1-212) x BD078483 (1-546)

Qy   127 ThrProMetLeuSerSerLysAlaSer 135
         |||||||||||||||||||||
Db   276 ACACCTATGCTCCTCCAAAGCTTCA 250

RESULT 74
CQ849043/c
LOCUS         CQ849043          546 bp     DNA       linear    PAT 19-AUG-2004
DEFINITION    Sequence 73 from Patent EP1445316.
ACCESSION     CQ849043
VERSION       CQ849043.1   GI:51470455
KEYWORDS      .
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominidae; Homo.
REFERENCE     1
AUTHORS       Ni,J., Young,P., Rosen,C.A., Florence,K.A., Ruben,S.M.,
              Carter,K.C., Feng,P., Endress,G.A., LaFleur,D.W., Shi,Y., Janat,F.,
              Duane,D.R., Greene,J.M., Ferrie,A.M. and Yu,G.L.
TITLE         Novel secreted protein
JOURNAL       Patent: EP 1445316-A 73 11-AUG-2004;
              Human Genome Sciences, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..546
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
Alignment Scores:
Pred. No.:               93.2          Length:          546
Score:                   9.00          Matches:         9
Percent Similarity:      100.0%        Conservative:    0
Best Local Similarity:   100.0%        Mismatches:      0
Query Match:             4.2%          Indels:          0
DB:                                    Gaps:            0

US-09-989-890-238 (1-212) x CQ849043 (1-546)

Qy   127 ThrProMetLeuSerSerLysAlaSer 135
         |||||||||||||||||||||
Db   276 ACACCTATGCTCCTCCAAAGCTTCA 250

RESULT 75
BV345691/c
LOCUS         BV345691          604 bp     DNA       linear    STS 27-JAN-2005
DEFINITION    S230P6615FF6.T0 Rottweiler Canis familiaris STS genomic, sequence
              tagged site.
ACCESSION     BV345691
VERSION       BV345691.1   GI:57599570
KEYWORDS      STS.
SOURCE        Canis familiaris (dog)
ORGANISM      Canis familiaris
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
              Canis.
REFERENCE     1  (bases 1 to 604)
AUTHORS       Lindblad-Toh,K.
TITLE         The genome sequence of Canis familiaris
JOURNAL       Unpublished (2004)
COMMENT       Contact: Kerstin Lindblad-Toh
```

Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersli@genome.wi.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 604
Protocol:
WGS-discovery (WGS):
Paired-end and low-coverage whole genome shotgun reads were generated
from 9 breeds
(German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
Retriever, English
Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
Water Dog ~100,000 each)
and five other canids (Chinese, Alaskan, Indian and Spanish Gray
Wolf as well as the
Californian Coyote).
The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
and SNP detection was
carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
485941 SNPs were
annotated with alleles from the boxer and the breed or canid from
which the particular
read came. The validation rate for these SNPs was estimated at
approximately 98%.
WGA-discovery (WGA) of Boxer/Poodle SNPs:
A second set of SNPs was generated using a similar methodology
except that the contigs
from the 1.5x poodle assembly (Kirkness 2003) were used instead of
WGS reads. Since this
sequence lacked base quality scores, arbitrary quality scores of
phred 40 were assigned
before the poodle sequence was placed uniquely on the CanFam1.0
boxer assembly and SNP
detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
with alleles from the
boxer and the poodle. The validation rate for these SNPs was
estimated at approximately TBD%.
Internal-WGA-discovery (I-WGA):
A third set of SNPs were discovered by comparing reads in the WGA
assembly. SNPs were
defined as mismatch positions that had a base quality of >= 30 on
both reads in a region
that aligned without gaps, and with at most one additional mismatch
in the ten flanking
bases. For each allele, at least one additional read had to confirm
it. 731476 SNPs were
annotated with alleles between the two boxer alleles. The
validation rate for these SNPs
was estimated at approximately TBD%.

FEATURES   Location/Qualifiers
     source   1..604
                /organism="Canis familiaris"
                /mol_type="genomic DNA"
                /strain="Rottweiler"
                /db_xref="taxon:9615"
                /map="+ 11 22-543 12395351-12394828"
                /clone_lib="Rottweiler"
     STS   <1..>604
ORIGIN

Alignment Scores:
Pred. No.:   102   Length: 604
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 10   Gaps: 0

US-09-989-890-238 (1-212) x BV345691 (1-604)

Qy   31 LeuSerArgProLeuSerProProPro 39
           |||||||||||||||||||||||||||
Db   493 CTCTCTCGGCGCCCTCTCCCCTCCTCCA 467

RESULT 76
LOCUS   CS074643   1119 bp   DNA   linear   PAT 05-MAY-2005
DEFINITION Sequence 69 from Patent WO2005030250.
ACCESSION   CS074643
VERSION   CS074643.1 GI:63091233
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tuereci,O., Sahin,U., Helftenbein,G. and Schlueter,V.
  TITLE   Identification of tumour-associated cell surface antigens for
            diagnosis and therapy
  JOURNAL   Patent: WO 2005030250-A 69 07-APR-2005;
            Ganymed Pharmaceuticals AG (DE)
FEATURES   Location/Qualifiers
     source   1..1119
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:   172   Length: 1119
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x CS074643 (1-1119)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
           |||||||||||||||||||||||||||
Db   801 TTGCCATCCTCTCGGCCTTGTTGTCC 827

RESULT 77
LOCUS   AY150443   1210 bp   mRNA   linear   PLN 23-SEP-2002
AY150443/c
DEFINITION Arabidopsis thaliana unknown protein (At1g51560) mRNA, complete
            cds.
ACCESSION   AY150443
VERSION   AY150443.1 GI:22297512
KEYWORDS   FLI_cDNA.
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 1210)
  AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
            Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
            Wu,H.C., Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J.,
            Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE   Arabidopsis Open Reading Frame (ORF) Clones
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1210)
  AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
            Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
            Wu,H.C., Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J.,
            Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-SEP-2002) Plant Gene Expression Center, 800 Buchanan

```
COMMENT     Street, Albany, CA 94710, USA
            Annotation is based on the January 2002 version of the Arabidopsis
            genome submitted to GenBank.
FEATURES             Location/Qualifiers
     source          1..1210
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="C104814"
                     /ecotype="Columbia"
                     /note="This clone is in pUNI 51."
     gene            1..1210
                     /gene="At1g51560"
     CDS             1..1179
                     /gene="At1g51560"
                     /codon_start=1
                     /evidence=experimental
                     /product="unknown protein"
                     /protein_id="AAN12985.1"
                     /db_xref="GI:23397513"
                     /translation="MDAAIFTSVYVCNIPKTKGAFNPNPPALSSSSCWLCNSQAKQI
                     IKLRIREGSNQGLLRVHALFNNEEASSESEDKNGFGLLPADIFSLPQEKFGSNVSGEK
                     DSENIIDVETSLAVPHGGGTRAGLFRTPISGGVQSATSAHGLPRPALAVRNLMEQARF
                     AHLCTVMSKMHRREGYPFGSLVDHFAPDMGHPIFSFSPLAIHTRNILAEPRCTLVVQ
                     IPGWSCLSNARVTLFGDVYPLPEEQQEWAHKQYMLKHHQGPSQQWGNFHYFRMQNISD
                     YFIGGFGTVAWINVNEYETLQPDKIAVDGGEQNLKELNAIFSKPLRELLSSEAELDD
                     AAIISIDSKGIDIRVRQGAQFKIQRLAFEESHGVETLEBARKSALWKVIEKGKLHNLQK
                     "
     3'UTR           1180..1210
                     /gene="At1g51560"
ORIGIN

Alignment Scores:
Pred. No.:             184            Length:        1210
Score:                 9.00           Matches:       9
Percent Similarity:    100.0%         Conservative:  0
Best Local Similarity: 100.0%         Mismatches:    0
Query Match:           4.2%           Indels:        0
DB:                    15             Gaps:          0

US-09-989-890-238 (1-212) x AY150443 (1-1210)

Qy  130 LeuSerSerLysAlaSerLeuCysIle 138
        |||||||||||||||||||||||||||
Db 1094 CTTTCCTCAAAGGCTAGCCTTTGTATC 1068

RESULT 78
AY508007/c
LOCUS       AY508007    1301 bp    mRNA    linear    PLN 01-JAN-2005
DEFINITION  Helianthus annuus DREB2 transcription factor (DREB2) mRNA, complete
            cds.
ACCESSION   AY508007
VERSION     AY508007.1    GI:46241281
KEYWORDS    .
SOURCE      Helianthus annuus (common sunflower)
  ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 1301)
  AUTHORS   Diaz-Martin,J., Almoguera,C. and Jordano,J.
  TITLE     A specific interaction between HaDREB2 and HaHSFA9 mediates
            transcriptional activation during embryo desiccation in sunflower
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1301)
  AUTHORS   Diaz-Martin,J., Almoguera,C. and Jordano,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-2003) Biologia Vegetal, IRNAS (CSIC), Av. Reina
            Mercedes, 10. Apartado 1052, Sevilla 41080, Spain
FEATURES             Location/Qualifiers
     source          1..1301
                     /organism="Helianthus annuus"
                     /mol_type="mRNA"
                     /cultivar="P113HS"
                     /db_xref="taxon:4232"
                     /dev_stage="developing embryo; 14 dpa"
     gene            1..1301
                     /gene="DREB2"
                     /note="HaDREB2"
     CDS             168..1112
                     /gene="DREB2"
                     /note="class A2 DREB transcription factor from the plant
                     AP2/ERF family; invoved in the developmental regulation of
                     small heat shock protein (sHSP) genes in embryos; shows
                     physical and specific genetic interaction with a second
                     transcription factor HSFA9 encoded by GenBank Accession
                     Number AY099451"
                     /codon_start=1
                     /product="DREB2 transcription factor"
                     /protein_id="AAS82861.1"
                     /db_xref="GI:46241282"
                     /translation="MALGFSDRNTEMSPYSQQSNTSAVPMDSSKKRKTRSRKEAPEKV
                     AATLAWEYNKANEDGKAKTRWAPAKGSKKGCMKGKGGPENSRTNFRGVRQRTWGKW
                     VAEIREPNRGRRLWLGTFGSAVEAALAYDEAARAMYGTSARLNLPNCRSKNDILPLLV
                     TNPASSCDSTTTCSYSEVDATHDSKPGPAVPFPSVKHBESVQVKHEPEIVAKEBHCVDN
                     NDLGFDIGDEMFDLEELLGAVNPEAGSGDGYDGRFVNGADANENMQMYMQPDPCV
                     SGQDYSPDFLQPGRPEDCSFTLEELGLGLDAELDLYSN"
     misc_feature    435..608
                     /gene="DREB2"
                     /note="DNA-binding domain; involved in protein-protein
                     interaction with HSFA9; Region: AP2 domain"
                     /evidence=experimental
     misc_feature    858..1097
                     /gene="DREB2"
                     /note="rich in acidic amino acids; required for functional
                     synergism with HSFA9; Region: transcription activation
                     domain"
                     /evidence=experimental
ORIGIN

Alignment Scores:
Pred. No.:             196            Length:        1301
Score:                 9.00           Matches:       9
Percent Similarity:    100.0%         Conservative:  0
Best Local Similarity: 100.0%         Mismatches:    0
Query Match:           4.2%           Indels:        0
DB:                    15             Gaps:          0

US-09-989-890-238 (1-212) x AY508007 (1-1301)

Qy   23 AlaLeuProSerSerLeuAlaLeuLeu 31
        |||||||||||||||||||||||||||
Db  355 GCTTTACCGTCTTCATTAGCCTTGTTA 329

RESULT 79
AY188374/c
LOCUS       AY188374    1327 bp    DNA    linear    PLN 06-JAN-2004
DEFINITION  Sporisorium sp. O-058.1d internal transcribed spacer 1, partial
            sequence; 5.8S ribosomal RNA, internal transcribed spacer 2,
            complete sequence; and 26S ribosomal RNA, partial sequence.
ACCESSION   AY188374
VERSION     AY188374.1    GI:28826190
KEYWORDS    .
SOURCE      Sporisorium sp. O-058.1d
  ORGANISM  Sporisorium sp. O-058.1d
            Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
            Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Sporisorium.
REFERENCE   1  (bases 1 to 1327)
  AUTHORS   Golonka,A.M.
  TITLE     Nectar-Inhabiting Yeasts in Silene latifolia and Co-flowering
            Plants
  JOURNAL   Thesis (2002) Duke University, Durham, NC, USA
REFERENCE   2  (bases 1 to 1327)
  AUTHORS   Golonka,A.M.
```

TITLE   Nectar-Inhabiting Yeasts in Silene latifolia and Co-flowering
           Plants
JOURNAL   Thesis (2002) Duke University, Durham, NC, USA
REFERENCE 3 (bases 1 to 1327)
AUTHORS   Golonka,A.M., Vilgalys,R. and Antonovics,J.
TITLE   Direct Submission
JOURNAL   Submitted (27-NOV-2002) Biology, Duke University, PO Box 90338,
           Durham, NC 27708, USA
FEATURES   Location/Qualifiers
     source   1..1327
                     /organism="Sporisorium sp. O-058.1d"
                     /mol_type="genomic DNA"
                     /strain="O"
                     /isolate="058.1d"
                     /isolation_source="nectar"
                     /db_xref="taxon:220537"
                     /note="species [or generic] identification based on nrDNA
                     sequence similarity"
                     maintained in culture"
     misc_RNA   <1..>1327
                     /note="contains internal transcribed spacer 1, 5.8S
                     ribosomal RNA, internal transcribed spacer 2, and 26S
                     ribosomal RNA"

ORIGIN

Alignment Scores:
Pred. No.:   200   Length: 1327
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 15   Gaps: 0

US-09-989-890-238 (1-212) x AY188374 (1-1327)

Qy   23 AlaLeuProSerSerLeuAlaLeuLeu 31
           |||||||||||||||||||||||||||
Db   774 GCTCTTCCCTCTTCACTCGCGTTACTA 748

RESULT 80
BC101317
BC101317
LOCUS   BC101317   1362 bp   mRNA   linear   PRI 02-AUG-2005
DEFINITION Homo sapiens cDNA clone MGC:120420 IMAGE:40024858, complete cds.
ACCESSION   BC101317
VERSION   BC101317.1 GI:71680579
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 1362)
AUTHORS   Strausberg,R.L., Feingold,B.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
CONSRTM   Mammalian Gene Collection Program Team
TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences

JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED   12477932
REFERENCE   2 (bases 1 to 1362)
AUTHORS   NIH MGC Project
CONSRTM   Direct Submission
TITLE   Submitted (01-AUG-2005) National Institutes of Health, Mammalian
JOURNAL   Gene Collection (MGC), Bethesda, MD 20892-2590, USA
            NIH-MGC Project URL: http://mgc.nci.nih.gov
REMARK   Contact: MGC help desk
COMMENT   Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Baylor Human Genome Sequencing Center
            cDNA Library Preparation: Baylor Human Genome Sequencing Center
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAM Plate: 7 Row: n Column: 8.
FEATURES   Location/Qualifiers
     source   1..1362
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:120420 IMAGE:40024858"
                     /tissue_type="PCR rescued clones"
                     /clone_lib="NIH_MGC_283"
                     /note="Vector: pCR-Blunt II-TOPO with reversed insert;
                     Clone identification sequence tag: NA"
     CDS   1..507
                     /codon_start=1
                     /product="Unknown (protein for MGC:120420)"
                     /protein_id="AAI01318.1"
                     /db_xref="GI:71680580"
                     /translation="MAFQDLLDQVGGLGRFQILQMVFLIMFNVIVYHQTQLENFAAFI
                     LDHRCWVHILDNTIPDNDPGTLSQDALLRISIPFDSNLRPEKCRRFVHPQWKLIHLN
                     GTFPNTSEPDTEPCVDGWVYDQQSFPSTIVTKVAGRVCSVAHYQQQTRRGLKGTYKSC
                     TQEWNEBC"

ORIGIN

Alignment Scores:
Pred. No.:   204   Length: 1362
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x BC101317 (1-1362)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
           |||||||||||||||||||||||||||
Db   1235 TTGCCATCCTCTCGGCCTTGTTGTCC 1261

RESULT 81
AR626918/c
LOCUS   AR626918   1377 bp   DNA   linear   PAT 14-FEB-2005
DEFINITION Sequence 8319 from patent US 6833447.
ACCESSION   AR626918
VERSION   AR626918.1 GI:59734648
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 1377)
AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
TITLE   Myxococcus xanthus genome sequences and uses thereof
JOURNAL   Patent: US 6833447-A 8319 21-DEC-2004;

```
FEATURES          Monsanto Technology, LLC; St. Louis, MO
source            Location/Qualifiers
                  1. .1377
                  /organism="unknown"
                  /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:        206          Length:         1377
Score:            9.00         Matches:        9
Percent Similarity: 100.0%     Conservative:   0
Best Local Similarity: 100.0%  Mismatches:     0
Query Match:      4.2%         Indels:         0
DB:                            Gaps:           0

US-09-989-890-238 (1-212) x AR626918 (1-1377)

Qy     56  SerAlaSerAlaAlaGlyIleAla  64
           ||||||||||||||||||||||||
Db   1289  AGTGCCAGCGCGCTGCGGAATCGCT  1263

RESULT 82
AY074356/c
LOCUS       AY074356          1436 bp    mRNA    linear    PLN 18-SEP-2002
DEFINITION  Arabidopsis thaliana unknown protein (At1g51560) mRNA, complete
            cds.
ACCESSION   AY074356
VERSION     AY074356.1  GI:18377804
KEYWORDS    FLI_cDNA.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1436)
AUTHORS     Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
TITLE       Arabidopsis Full Length cDNA Clones
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 1436)
AUTHORS     Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,
            Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,
            Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
TITLE       Direct Submission
JOURNAL     Submitted (11-JAN-2002) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Banh,J.,
            Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M.,
            Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C.,
            Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Chen,H.,
            Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P.,
            Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
```

```
            contributed equally to this work as PIs.

            Annotation is based on the January 2002 version of the Arabidopsis
            genome submitted to GenBank.
FEATURES          Location/Qualifiers
source            1. .1436
                  /organism="Arabidopsis thaliana"
                  /mol_type="mRNA"
                  /db_xref="taxon:3702"
                  /chromosome="1"
                  /clone="RAFL08-11-G16 (R11062)"
                  /ecotype="Columbia"
                  /note="This clone is in a modified pBluescript vector
                  (PlC-1) as a BamHI/XhoI insert."
gene              1. .1436
                  /gene="At1g51560"
5'UTR             1. .78
                  /gene="At1g51560"
CDS               79. .1257
                  /gene="At1g51560"
                  /codon_start=1
                  /evidence=experimental
                  /product="unknown protein"
                  /protein_id="AAL67052.1"
                  /db_xref="GI:18377805"
                  /translation="MDAAIFTSVYVCNIPKTKKAFPNPNPPALSSSSCWLCNSQAKQI
                  IKLRIREGSNQGLLRVHALFNNEEASSESEDKNGFGLLPADIFSLPQEKFGSNVSGEK
                  DSENIIDVETSLAVPHGGGTRAGLYRTPISGGVQSATSAHGLRPPALAVRNLMEQARF
                  AHLCTVMSKMHRREGYPFGSLVDFAPDMGHPIFSFSPLAIHTRNILAEPRCTLVVQ
                  IPGWSCLSNARVTLFGDVYPLPEEQQEWAHKQYMLKHHQGPSCQWGNFHYFRMQNISD
                  IYFIGGFTVAWINVNEYETLQPDKIAVDGGEONLKELNAIFSKPLRELLSSEABLDD
                  AAIISIDSKGIDIRVRQGAQFKIQRLAFEESHGVETLBEAKSALWKVIEKGKLHNLQK
                  "
misc_difference   458
                  /gene="At1g51560"
                  /note="compared to genomic sequence resulting in an amino
                  acid sequence difference"
                  /replace="t"
3'UTR             1258. .1436
                  /gene="At1g51560"
misc_difference   1419. .1420
                  /gene="At1g51560"
                  /note="not in genomic sequence"
ORIGIN

Alignment Scores:
Pred. No.:        214          Length:         1436
Score:            9.00         Matches:        9
Percent Similarity: 100.0%     Conservative:   0
Best Local Similarity: 100.0%  Mismatches:     0
Query Match:      4.2%         Indels:         0
DB:                            Gaps:           15

US-09-989-890-238 (1-212) x AY074356 (1-1436)

Qy    130  LeuSerSerLysAlaSerLeuCysIle  138
           |||||||||||||||||||||||||||
Db   1172  CTTTCCTCAAAGGCTAGCCTTTGTATC  1146

RESULT 83
AF153341/c
LOCUS       AF153341          1448 bp    DNA    linear    PRI 24-AUG-2001
DEFINITION  Homo sapiens winged helix/forkhead transcription factor (HFH1)
            gene, complete cds.
ACCESSION   AF153341
VERSION     AF153341.1  GI:8489092
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1448)
```

AUTHORS   Hong,H.K., Noveroske,J.K., Headon,D.J., Liu,T., Sy,M.S.,
           Justice,M.J. and Chakravarti,A.
TITLE   The winged helix/forkhead transcription factor Foxq1 regulates
           differentiation of hair in satin mice
JOURNAL   Genesis 29 (4), 163-171 (2001)
PUBMED   11309849
REFERENCE 2 (bases 1 to 1448)
AUTHORS   Hong,H.-K., Noveroske,J.K., Justice,M.J. and Chakravarti,A.
TITLE   Direct Submission
JOURNAL   Submitted (21-MAY-1999) Genetics, Case Western Reserve University,
           10900 Euclid Avenue, Cleveland, OH 44106-4955, USA
FEATURES   Location/Qualifiers
     source   1..1448
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p25"
     gene   <233..>1441
                     /gene="HFH1"
     mRNA   <233..>1441
                     /gene="HFH1"
                     /product="winged helix/forkhead transcription factor"
     CDS   233..1441
                     /gene="HFH1"
                     /codon_start=1
                     /product="winged helix/forkhead transcription factor"
                     /protein_id="AAF75586.1"
                     /db_xref="GI:8489093"
                     /translation="MKLEVFVPRAAHGDKQGSDLEGAGGSDAPSPLSAAGDDSLGSDG
                     DCAAKPSAGGARDTOGDGEOSAGGGPGAEEAIPAAAAAVVAEGAEAGAGPGAGGA
                     GSGEGARSKPYTRRPKPPYSYIALIAMAIRDSAGGRLTLAEINEYLMGKFPFFRGSYT
                     GWRNSVRHNLSLNDCFVVKLRDPSRPWGKDNTWMLNPNSEYTFADGVFRRRRKRLSHR
                     APVPADGLRPEEAPGLPAAPPAPAAPASPRMRSPARQEERASPAGKPSSSFAIDSIL
                     RKPFRSRRLRDTAPGTTIQWGAAPCPPLPAFPALLPAAPCRALLPLCAYGAGEPARLG
                     AREAEVPPTAPPLLLAPLPAAAPAKPLRGPAAGGAHLYCPLRLPAALQAALVRRPGPH
                     LSYPVETLLA"

ORIGIN

Alignment Scores:
Pred. No.:   215   Length: 1448
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x AF153341 (1-1448)

Qy   59 AlaAlaAlaGlyIleAlaSerSerAla 67
         |||||||||||||||||||||||||||
Db   474 GCTGCTGCCGGGATCGCCTCCTCCGCG 448

RESULT 84
AX685724
LOCUS   AX685724   1644 bp   DNA   linear   PAT 29-MAR-2003
DEFINITION Sequence 1 from Patent WO02079252.
ACCESSION   AX685724
VERSION   AX685724.1 GI:29371736
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
AUTHORS   Merkulov,G. and Beasley,E.M.
TITLE   Isolated human transporter protein, nucleic acid molecules encoding
            human transporter protein, and uses thereof
JOURNAL   Patent: WO 02079252-A1 10-OCT-2002;
            PE Corporation (NY) (US)
FEATURES   Location/Qualifiers
     source   1..1644

                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.:   240   Length: 1644
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x AX685724 (1-1644)

Qy   24 LeuProSerLeuAlaLeuLeuSer 32
         ||||||||||||||||||||||||
Db 1511 TTGCCATCCTCTCTGGCCTTGTTGTCC 1537

RESULT 85
AY437532
LOCUS   AY437532   1644 bp   mRNA   linear   PRI 12-NOV-2004
DEFINITION Homo sapiens putative UST1-like organic anion transporter mRNA,
            complete cds.
ACCESSION   AY437532
VERSION   AY437532.1 GI:40287954
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 1644)
AUTHORS   Eraly,S.A., Monte,J.C. and Nigam,S.K.
TITLE   Novel slc22 transporter homologs in fly, worm, and human clarify
            the phylogeny of organic anion and cation transporters
JOURNAL   (er) Physiol. Genomics 18 (1), 12-24 (2004)
PUBMED   15054140
REFERENCE   2 (bases 1 to 1644)
AUTHORS   Monte,J.C., Nigam,S.K. and Eraly,S.A.
TITLE   Direct Submission
JOURNAL   Submitted (13-OCT-2003) Medicine, University of California, San
            Diego, 9500 Gilman Drive, La Jolla, CA 92093-0693, USA
FEATURES   Location/Qualifiers
     source   1..1644
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="kidney"
     CDS   1..1644
                     /note="UST6"
                     /codon_start=1
                     /product="putative UST1-like organic anion transporter"
                     /protein_id="AAR84082.1"
                     /db_xref="GI:40287955"
                     /translation="MAFQDLLDQVGGLRGFQILQMVFLIMFNVIVYHQTQLENPAAFI
                     LDHRCWVHILDNDTIPDNDPGTLSQDALLRISIPFDSNLRPEKCRRPFVHPQWKLIHLN
                     GTFPNTSEPDTBPCVDGWVYDOSSFPSTIVTKWDLVCRSQPLNSVAKFLFMAGMMVGG
                     NLYGHLSDRFGRKFVLRWSYLQLAIVGTCAAFAPTILVYCSLRFLAGAATFSIIVNTV
                     LLIVEWITHQFCAMALTLTLCAASIGHTLGSLAFVIRDQCILQLVMSAPCFVFLFS
                     RWLABSARWLIINNKPEEGLKELRKAAHRNGMKNAEDLLTMBVLKSTMKQBLERAQKK
                     HSICELLRIPNICKRICFLSFVRFASTIPFWGLTLHLQHGNNVFLLQFLFGAVTLLA
                     NCVAPWALNHMSRLSQMLMFPLIATCLLAIIFVPQEMQTLRVVLATLGVGAASLGIT
                     CSTAQBNELIPSIIRGRATGITGNFANIGGALASLVMLSIYSRPLPWIIYGVFAILS
                     GLVVLLLPETRNQPLLDSIQDVENEGVNSLAAPQRSSVL"

ORIGIN

Alignment Scores:
Pred. No.:   240   Length: 1644
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0

DB: Gaps: 0

US-09-989-890-238 (1-212) x AY437532 (1-1644)

Qy   24 LeuProSerSerLeuAlaAlaLeuLeuSer 32
               ||||||||||||||||||||||||||||||
Db   1511 TTGCCATCCTCTCTGGCCTTGTTGTCC 1537

RESULT 86
AR627633/c
LOCUS   AR627633   1716 bp   DNA   linear   PAT 14-FEB-2005
DEFINITION Sequence 9034 from patent US 6833447.
ACCESSION   AR627633
VERSION   AR627633.1 GI:59735932
KEYWORDS   Unknown.
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 1716)
  AUTHORS   Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
  TITLE   Myxococcus xanthus genome sequences and uses thereof
  JOURNAL   Patent: US 6833447-A 9034 21-DEC-2004;
            Monsanto Technology, LLC; St. Louis, MO
FEATURES   Location/Qualifiers
     source   1..1716
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:   249   Length: 1716
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x AR627633 (1-1716)

Qy   50 SerGlyAlaGlyLeuProSerAlaSer 58
               ||||||||||||||||||||||||||||
Db   1342 TCAGGTGCTGGCCTTCCTTCAGCGAGT 1316

RESULT 87
D89272/c
LOCUS   D89272   1761 bp   mRNA   linear   PLN 13-MAR-1998
DEFINITION Schizosaccharomyces pombe mRNA, partial cds, clone: SY 1780.
ACCESSION   D89272
VERSION   D89272.1 GI:1749751
KEYWORDS   .
SOURCE   Schizosaccharomyces pombe (fission yeast)
  ORGANISM Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1 (sites)
  AUTHORS   Yoshioka,S., Kato,K., Nakai,K., Okayama,H. and Nojima,H.
  TITLE   Identification of open reading frames in Schizosaccharomyces pombe
            cDNAs
  JOURNAL   DNA Res. 4 (6), 363-369 (1997)
   PUBMED   9501991
REFERENCE   2 (bases 1 to 1761)
  AUTHORS   Yoshioka,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-NOV-1996) Sachiyo Yoshioka, Tsukita Cell Axis Project
            ERATO JST, Kyoto Research Park; 17 Chudouji Minamimachi,
            Shimokyo-ku, Kyoto, Kyoto 600, Japan
            (E-mail:syoshio@cell.tsukita.jst.go.jp, Tel:+81-75-315-7913,
            Fax:+81-75-315-6420)
FEATURES   Location/Qualifiers
     source   1..1761
                     /organism="Schizosaccharomyces pombe"
                     /mol_type="mRNA"

CDS

                     /strain="PR745"
                     /db_xref="taxon:4896"
                     /clone="SY1780"
                     /clone_lib="library of H. Nojima"
                     <1..1676
                     /note="similar to Saccharomyces cerevisiae T-complex
                     protein 1,theta subunit, SWISS-PROT Accession Number
                     P47079"
                     /codon_start=3
                     /protein_id="BAA13933.1"
                     /db_xref="GI:1749752"
                     /translation="FYQDIKAFVNKMALRVPKASGPQLFREGYRIMQGVEDAVIRNCN
                     AIRELSEITRTSLGPNGKNKIVVNHLQQTFLTNDAATIIRELEVIHPAAKLVDATQQ
                     QENELGDAANFVIVFSGALLLKAENMIRMGLTPLEIAKGYEMALSHTMEVLESICADK
                     IETVESEKELIKAIRTCISSKQYGNEDFLSDLVAKAILTVLPKDPSKFNVDNIRVVKI
                     MGSSLYNSQVVKGMVFPRPEPEGTVTRSKEAKVAVFSCPLDISQETERKGTVLLHNAQEM
                     LDFSKGENLIESHIKSIYDAGVRVVVTSGNVNDLVLHYLNRFEILVIRVPSKFELRR
                     LCRVVGATPLARMGVPMPEEMGSVDVVETIEIGGDRVTVFRQVEDITRTATIVLRGAT
                     KTYLDDLERAIDDGVNIVKALVKDNRLIFGAGASDMOLCIRLISVGEKTPGIYQHAIK
                     QYGEAFEVVPRTISENAGLDPTDVISKLYAAHHKENGSIGVDVECENDGTLDAKEAG
                     IFDVLLAKKSAIRLATETVLTVLNVDQVVMSKPAGGPKPPGPNPHWDDD"
ORIGIN

Alignment Scores:
Pred. No.:   254   Length: 1761
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 15   Gaps: 0

US-09-989-890-238 (1-212) x D89272 (1-1761)

Qy   175 ValLeuValAlaProArgSerThrIle 183
               |||||||||||||||||||||||||||
Db   1183 GTCTTAGTAGCCCCTCGAGCACGATT 1157

RESULT 88
AX687001
LOCUS   AX687001   1818 bp   DNA   linear   PAT 31-MAR-2003
DEFINITION Sequence 61 from Patent WO0204520.
ACCESSION   AX687001
VERSION   AX687001.1 GI:29409522
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Raumann,B.E., Thornton,M., Ding,L., Yue,H., Tang,Y.T., Harland,L.,
            Burford,N., Greene,B.D., Sanjanwala,M., Baughn,M.R., Yao,M.,
            Yang,J., Patterson,C., Gandhi,A., Hafalia,A., Tribouley,C.,
            Walia,N., Au-Young,J., Walsh,R.T., Ramkumar,J., Lu,Y., Lu,D.A.,
            Azimzai,Y., Lal,P., Elliott,V., Nguyen,D., Xu,Y., Seilhamer,J.,
            Borowsky,M., Kahn,F.A., Kearney,L., Thangavelu,K., Das,D. and
            Policky,J.
  TITLE   Transporters and ion channels
  JOURNAL   Patent: WO 0204520-A 61 17-JAN-2002;
            Incyte Genomics, Inc. (US)
FEATURES   Location/Qualifiers
     source   1..1818
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 16882?CB1"
ORIGIN

Alignment Scores:
Pred. No.:   261   Length: 1818
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0

```
Query Match:        4.2%        Indels:     0
DB:                 6           Gaps:       0

US-09-989-890-238 (1-212) x AX687001 (1-1818)

Qy      24 LeuProSerSerLeuAlaLeuLeuSer 32
           ||||||||||||||||||||||||||
Db    1685 TTGCCATCCTCTCTGGCCTTGTGTCC 1711

RESULT 89
CQ876112
LOCUS       CQ876112              1912 bp    DNA         linear   PAT 04-OCT-2004
DEFINITION  Sequence 1 from Patent WO2004081041.
ACCESSION   CQ876112
VERSION     CQ876112.1  GI:53789716
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Golz,S., Brueggemeier,U. and Geerts,A.
  TITLE     Organic anion transporting (oat)-like protein ust3-like1 and
            useesthereof
  JOURNAL   Patent: WO 2004081041-A 1 23-SEP-2004;
            Bayer HealthCare AG (DE)
FEATURES             Location/Qualifiers
     source          1..1912
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

US-09-989-890-238 (1-212) x CQ876112 (1-1912)

Qy      24 LeuProSerSerLeuAlaLeuLeuSer 32
           ||||||||||||||||||||||||||
Db    1589 TTGCCATCCTCTCTGGCCTTGTGTCC 1615

RESULT 90
AX687002
LOCUS       AX687002              2245 bp    DNA         linear   PAT 31-MAR-2003
DEFINITION  Sequence 62 from Patent WO0204520.
ACCESSION   AX687002
VERSION     AX687002.1  GI:29409523
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Raumann,B.B., Thornton,M., Ding,L., Yue,H., Tang,Y.T., Harland,L.,
            Burford,N., Greene,B.D., Sanjanwala,M., Baughn,M.R., Yao,M.,
            Yang,J., Patterson,C., Gandhi,A., Hafalia,A., Tribouley,C.,
            Walsh,N., Au-Young,J., Walsh,R.T., Ramkumar,J., Lu,Y., Lu,D.A.,
            Azimzai,Y., Lal,P., Elliott,V., Nguyen,D., Xu,Y., Seilhamer,J.,
            Borowsky,M., Kahn,F.A., Kearney,L., Thangavelu,K., Das,D. and
            Policky,J.
  TITLE     Transporters and ion channels
  JOURNAL   Patent: WO 0204520-A 62 17-JAN-2002;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2245
```

```
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 7472734CB1"
ORIGIN

Alignment Scores:
Pred. No.:              313              Length:            2245
Score:                  9.00             Matches:           9
Percent Similarity:     100.0%           Conservative:      0
Best Local Similarity:  100.0%           Mismatches:        0
Query Match:            4.2%             Indels:            0
DB:                     6                Gaps:              0

US-09-989-890-238 (1-212) x AX687002 (1-2245)

Qy      24 LeuProSerSerLeuAlaLeuLeuSer 32
           ||||||||||||||||||||||||||
Db    2112 TTGCCATCCTCTCTGGCCTTGTGTCC 2138

RESULT 91
BC053850/c
LOCUS       BC053850              2322 bp    mRNA        linear   PRI 08-MAR-2005
DEFINITION  Homo sapiens forkhead box Q1, mRNA (cDNA clone MGC:61699
            IMAGE:6010412), complete cds.
ACCESSION   BC053850
VERSION     BC053850.1  GI:31753144
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2322)
  AUTHORS   Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,
            Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, Zeeberg
            B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T,
            Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin
            GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL,
            Scheetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P,
            Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy
            SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH,
            Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW,
            Villalon DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J,
            Helton E, Ketteman M, Madan A, Rodrigues S, Sanchez A, Whiting M,
            Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW,
            Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J,
            Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U,
            Smailus DE, Schnerch A, Schein JE, Jones SJ and Marra MA.
  CONSRTM   Mammalian Gene Collection Program Team
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED    12477932
REFERENCE   2  (bases 1 to 2322)
  AUTHORS   Director MGC Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
```

```
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,       AUTHORS   Smith,T.P., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,                       Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,                 Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,              Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F.,
            McDowell,J., Pearson,R., Stanripop,S., Thomas,P.J., Touchman,J.W.,              Quackenbush,J. and Keele,J.W.
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,       TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            Young,A., Zhang,L.-H. and Green,E.D.                                            libraries and construction of a gene index for cattle
                                                                                 JOURNAL   Genome Res. 11 (4), 626-630 (2001)
            Clone distribution: MGC clone distribution information can be found   PUBMED    11282978
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov      REFERENCE 2  (bases 1 to 2386)
            Series: IRAK Plate: 115 Row: i Column: 14                             AUTHORS   Harhay,G.P., Sonstegard,T.S., Van Tassell,C.P., Clawson,M.L.,
            This clone was selected for full length sequencing because it                   Heaton,M.P., Keele,J.W., Snelling,W.M., Weidmann,R.T. and
            passed the following selection criteria: Hexamer frequency ORF                  Smith,T.P.L.
            analysis, GenomeScan gene prediction.                                 TITLE     Sequencing and analysis of Bos taurus full-length insert cDNA
FEATURES             Location/Qualifiers                                                    clones
     source          1..2322                                                      JOURNAL   Unpublished
                     /organism="Homo sapiens"                                     REFERENCE 3  (bases 1 to 2386)
                     /mol_type="mRNA"                                             AUTHORS   Harhay,G.P., Sonstegard,T.S., Van Tassell,C.P., Clawson,M.L.,
                     /db_xref="taxon:9606"                                                   Heaton,M.P., Keele,J.W., Snelling,W.M., Weidmann,R.T. and
                     /clone="MGC:61699 IMAGE:6010412"                                        Smith,T.P.L.
                     /tissue_type="Lung, large cell carcinoma"                    TITLE     Direct Submission
                     /clone_lib="NIH_MGC_68"                                      JOURNAL   Submitted (17-MAR-2005) Molecular Genetic Research Unit, USDA-ARS
                     /lab_host="DH10B"                                                       U.S. Meat Animal Research Center, P.O. Box 166, Spur 18D, Clay
                     /note="Vector: pCMV-SPORT6"                                             Center, NE 68933
     gene            1..2322                                                      COMMENT   Contact: Gregory P. Harhay at harhay@email.marc.usda.gov.
                     /gene="FOXQ1"                                                FEATURES             Location/Qualifiers
                     /note="synonym: HFH1"                                             source          1..2386
                     /db_xref="GeneID:94234"                                                           /organism="Bos taurus"
     CDS             266..1477                                                                         /mol_type="mRNA"
                     /gene="FOXQ1"                                                                     /db_xref="taxon:9913"
                     /codon_start=1                                                                    /clone="3BOV59119; plate 59 row I column 19"
                     /product="forkhead box Q1"                                                        /tissue_type="pooled"
                     /protein_id="AAH53850.1"                                                          /clone_lib="3BOV"
                     /db_xref="GI:31753145"                                                            /lab_host="Escherichia coli DH10B"
                     /db_xref="GeneID:94234"                                                           /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     /translation="MKLEVFVPRAAHGDKQGSDLEGAGGSDAPSPLSAAGDDSLGSDG                             Library made from pooled tissue from marrow, alveolar
            DCAANSPAAGGAARDPPGDGEQSAGGGPGAEEAIPAAAAAAVVAEGAEAGAAGPGAGG                                  macrophage, ovary, fetal semitendonosus muscle, and fetal
            AGSGGEGARSKPYTRRPKPPYSYIALIAMAIRDSAGGRLTLAEINEYLMGKFPFFRGSY                                 longissimus muscle."
            TGWRNSVRHNLSLNDCFVKVLRDPSRPWGKDNYWMLNPNSEYTFADGVFRRRRKRLSH                      gene            1..2386
            RAPVPAPGLRPEEAPGLRPAAPPPAPRAPASPRMRSPARQEERASPAGPGPAKGKFSSSFAIDSI                            /gene="PISD"
            LRKPFRSRRLRDTAPGTTLQWGAAPCPPLPAPPALLPAAAPCRALLPLCAYGAGEPARL                                 /note="putative, similar to GeneID: 23761"
            GAREAEVPPTAPPLLLAPLPAAAPAKPLRGPAAGAHLYCPLRLPAALQAASVRRPGP         CDS             137..1387
            HLPYPVETLLA"                                                                               /gene="PISD"
                                                                                                      /codon_start=1
                                                                                                      /product="phosphatidylserine decarboxylase"
ORIGIN                                                                                                 /protein_id="AAX46472.1"
Alignment Scores:                                                                                     /db_xref="GI:61553873"
Pred. No.:              323            Length:            2322                                         /translation="MPGKSTRPLPAPRPCCAPCPFWSPRLAGMQGTGSTRSTGSESWR
Score:                  9.00           Matches:           9                                   SWAWTSRPSLLVTGRLHFPQLALRRRLGQLSCMSKPALKLRSWPLTVLYLPLPGLALR
Percent Similarity:    100.0%          Conservative:      0                                   PLSRVGWRPVSRVALYKSVPTRLLSRAWGRLNQVELPHWLRRPVYSLYIWTFGVNMKE
Best Local Similarity: 100.0%          Mismatches:        0                                   AAVEDLHYRNLSEFFRRKLKPQARPVCGLHSVISPSDGKILNFGQVKNCEVEBQVKGV
Query Match:           4.2%            Indels:            0                                   TYSLESFLGPRTPSEDLPFPPATPHSSFRSQLVTREGNELYHCVIYLAPGDYHCFHSP
DB:                                    Gaps:              8                                   TDWTVSHRRHFPGSLMSVNPGMARWIKELFCHNERVVLTGDWKHGFFSLTAVGATNVG
                                                                                             SIRIYFDRDLHTNSPRYSKGSYNDFSFVTHANKEGIPMRKGEHLGEFNLGSTIVLIFE
US-09-989-890-238 (1-212) x BC053850 (1-2322)                                                APKDFNFRLQAGQKIRFGEALGSL"

Qy    59 AlaAlaAlaGlyIleAlaSerSerAla 67                                 ORIGIN
         |||||||||||||||||||||||||||                                   Alignment Scores:
Db   510 GCTGCTGCCGGGATCGCCTCCTCCGCG 484                               Pred. No.:              330            Length:            2386
                                                                       Score:                  9.00           Matches:           9
                                                                       Percent Similarity:    100.0%          Conservative:      0
RESULT 92                                                              Best Local Similarity: 100.0%          Mismatches:        0
BT021625/c                                                             Query Match:           4.2%            Indels:            0
LOCUS       BT021625                2386 bp    mRNA       linear       DB:                                    Gaps:              0
            MAM 17-MAR-2005
DEFINITION  Bos taurus phosphatidylserine decarboxylase (PISD), mRNA, complete   US-09-989-890-238 (1-212) x BT021625 (1-2386)
            cds
ACCESSION   BT021625                                                  Qy    47 GlyCysGlySerGlyAlaGlyLeuPro 55
VERSION     BT021625.1  GI:61553872                                            |||||||||||||||||||||||||||
KEYWORDS    FLI_cDNA.                                                 Db   650 GGTTGCGGTAGTGGTGCAGGTCTTCCA 624
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus                                                RESULT 93
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;    BC085632
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
  REFERENCE 1  (bases 1 to 2386)
```

```
LOCUS       BC085632              2417 bp    mRNA    linear   VRT 06-JUL-2005
DEFINITION  Danio rerio zgc:92087, mRNA (cDNA clone MGC:92087 IMAGE:7047030),
            complete cds.
ACCESSION   BC085632
VERSION     BC085632.1  GI:5525068l
KEYWORDS    MGC.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 2417)
AUTHORS     Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
CONSRTM     Mammalian Gene Collection Program Team
TITLE       Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED      12477932
REFERENCE   2  (bases 1 to 2417)
CONSRTM     NIH MGC Project
TITLE       Direct Submission
JOURNAL     Submitted (01-NOV-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Bethesda, MD 20892-2590, USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Len Zon, Harvard
            cDNA Library Preparation: Open Biosystems
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site:     http://www-shgc.stanford.edu
            Contact:    (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 179 Row: c Column: 19
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 55250681.
FEATURES             Location/Qualifiers
     source          1..2417
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="MGC:92087 IMAGE:7047030"
                     /tissue_type="Whole, adult zebrafish"
                     /clone_lib="NIH_ZGC_10"
                     /lab_host="DH10B"
                     /note="Vector: pExpress1"
     gene            1..2417
                     /gene="zgc:92087"
                     /db_xref="GeneID:492333"
                     /db_xref="ZFIN:ZDB-GENE-041114-8"
     CDS             35..2032
                     /gene="zgc:92087"
                     /codon_start=1
                     /product="zgc:92087"
                     /protein_id="AAH85632.1"
                     /db_xref="GI:55250682"
                     /db_xref="GeneID:492333"
                     /db_xref="ZFIN:ZDB-GENE-041114-8"
                     /translation="MTDLQSKYSKLAQEYSKLRAQNQVLKKAVVDEQSCNSLKDELK
                     QKEQSLRRVEQEMDSLSFRNQLMKRVELLQEELLLSESKSKKSRSKGDSPSQVSLQA
                     QSVFDEDLHKKIQENERLHIQWLEAQEOHQQQRAQLSRLQOLQEEAQEHQAQLEELS
                     CRRAHTIHTLQEDXATLEVKLQTLEREARDCRVRTEECCQQLRKYQSEVSSLQKHSSS
                     VIQEKVPFNDTQLSDYNSLNVPAHNRRHQLRAREVAAQALVFLQNLVSALLNFHSYSE
                     QRVQIYPRDSSIETISAVNQKFSQYLHENASYLRQLEBDLLQLHQSITEDLPAATQKF
                     CTTNECLLSSLASLTSSSGKMATFFSNSLDFLTSCAGYSPSGALLKPLQADSVMQSKR
                     RAAYISSVRQARAESVPYGEALANRHILTSSTESREGLMQQVLQSQQKISRLEQEKE
                     HWLLEAQLAQVRLQKESARIAQLRAQVCVSAPQSQLCASAAESPVCVSAAESQVCVSA
                     AETPSEPALADAPEPLQDTSVVQLVISRSCSESSADQDSREQLIKTHYMSRVSELTTQ
                     LQICDSKAVHFHAECRAVAKRLAMAERSRDTLGEELKLANQNITRLQDELSTTKRSYE
                     DQLSMMSDHLCSMNETLSQQRETIDTLKLSAKGNAKKNKSR"
ORIGIN

Alignment Scores:
Pred. No.:                334          Length:              2417
Score:                   9.00          Matches:                9
Percent Similarity:     100.0%         Conservative:           0
Best Local Similarity:  100.0%         Mismatches:             0
Query Match:              4.2%         Indels:                 0
DB:                         5          Gaps:                   0

US-09-989-890-238 (1-212) x BC085632 (1-2417)

Qy     47 GlyCysGlySerGlyAlaGlyLeuPro 55
          |||||||||||||||||||||||||||
Db    790 GGTTGCGGCTCAGGCGCTGGTCTTCCT 816

RESULT 94
AK073176
LOCUS       AK073176              2576 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:J033004J13, full
            insert sequence.
ACCESSION   AK073176
VERSION     AK073176.1  GI:32983199
KEYWORDS    FLI_CDNA; CAP trapper.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
AUTHORS     The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team.,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group; Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN.,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
TITLE       Collection, mapping, and annotation of over 28,000 cDNA clones from
            japonica rice
JOURNAL     Science 301 (5631), 376-379 (2003)
PUBMED      12869764
REFERENCE   2  (bases 1 to 2576)
AUTHORS     Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
```

```
Qy      33  ArgProLeuSerProProProAlaAla 41
            |||||||||||||||||||||||||
Db     199  CGCCCGCTTCTCCGCCACCGGCAGCC 225
```

RESULT 95
HUMAGCGB
LOCUS   HUMAGCGB   2893 bp   mRNA   linear   PRI 12-JUN-1993
DEFINITION   Human chromosome 3p21.1 gene sequence, complete cds.
ACCESSION   L13434
VERSION   L13434.1 GI:291843
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Hominidae; Homo.
REFERENCE   1 (bases 1 to 2893)
  AUTHORS   Shridhar,V., Kamat,A.K., Golembieski,W., Smith,S.E.,
                Siegfried,J.M., Hunt,J.D., Miller,O.J., Wozniak,A. and Smith,D.I.
  TITLE   Identificaiton of new genes from human chromosomal band 3P21.1 and
                their levels of expression in lung cancer cell lines
  JOURNAL   Unpublished (1993)
COMMENT   Original source text: Homo sapiens adenogastric carcinoma, gastric
                mucosa cDNA to mRNA.
FEATURES   Location/Qualifiers
     source   1..2893
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="AGS"
                     /tissue_type="adenogastric carcinoma, gastric mucosa"
     CDS   1627..2397
                     /codon_start=1
                     /protein_id="AAA72367.1"
                     /db_xref="GI:291844"
                     /translation="METVVCPRPWEERRKRRSLSSDRGRTTHSPYEBRRSRTKGSGQQ
                     SERGSDRTPERSRKENHSSEGTKESSSNSLSNSRHGAEERGHHHHHEAADSSHGKKA
                     RDSERNHRTTEAEPKPLEEPKHETKKLKNLSEYAQTLQLGWNGLLVLKNSCFPTSMHI
                     LEGDQGVISSLLKDHTSGSKUTQLKIAQRLRLDQPKLDEDTRRIKQGSPNGYAVLLAT
                     QATPSGLGTEGMPTVEPGLQRRLLRNLVSYLKQKQAAG"
ORIGIN

Alignment Scores:
Pred. No.:   389   Length: 2893
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x HUMAGCGB (1-2893)
```

```
Qy      37  ProProProAlaAlaCysSerGlyAsp 45
            |||||||||||||||||||||||||
Db      57  CCGCCGCCCGCCGCCGTGTTCCGGAGAC 83
```

RESULT 96
AK065775
LOCUS   AK065775   3353 bp   mRNA   linear   PLN 24-JUL-2003
DEFINITION   Oryza sativa (japonica cultivar-group) cDNA clone:J013038N04, full
                insert sequence.
ACCESSION   AK065775
VERSION   AK065775.1 GI:32975793
KEYWORDS   FLI_CDNA; CAP trapper.
SOURCE   Oryza sativa (japonica cultivar-group)
  ORGANISM   Oryza sativa (japonica cultivar-group)
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
                Agrobiological Sciences Rice Full-Length cDNA Project Team:,
                Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
                Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
                Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
                Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
                Imotani,K., Ishikawa,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
                Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
                Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
                Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
                Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
                Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
                Mizuno,K., Murakami,K., Murata,M., Nagata,J., Nakamura,M.,
                Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K.,
                Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
                Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
                Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
                Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
                Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
                Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
                Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
                Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
                Yoshimura,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
                Agrobiological Sciences, Department of Molecular Genetics, Head of
                Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
                305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
                Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT   This clone is one of the 28K full-length cDNA clones from japonica
                rice.
                URL : http://cdna01.dna.affrc.go.jp/cDNA/
                NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,
                Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
                Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., and
                Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and
                Yamamoto,M.
                FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y.,
                Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
                Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
                Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S.,
                Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
                Yoshimura,A., Matsubara,K. and Murakami,K.
                Genome Exploration Research Group in Riken Genomic Sciences Center
                and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
                Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
                Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
                Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
                Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
                Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
                Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
                Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,
                Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
                Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
                Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
                Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
                Yasunishi,A. and Hayashizaki,Y.
FEATURES   Location/Qualifiers
     source   1..2576
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="J013004J13"
ORIGIN

Alignment Scores:
Pred. No.:   353   Length: 2576
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 15   Gaps: 0

US-09-989-890-238 (1-212) x AK073176 (1-2576)
```

FEATURES          Yasunishi,A. and Hayashizaki,Y.
source            Location/Qualifiers
                  1..3353
                  /organism="Oryza sativa (japonica cultivar-group)"
                  /mol_type="mRNA"
                  /cultivar="Nipponbare"
                  /db_xref="taxon:39947"
                  /clone="J013038N04"

ORIGIN

Alignment Scores:
Pred. No.:                442          Length:                3353
Score:                    9.00         Matches:               9
Percent Similarity:       100.0%       Conservative:          0
Best Local Similarity:    100.0%       Mismatches:            0
Query Match:              4.2%         Indels:                0
DB:                       15           Gaps:                  0

US-09-989-890-238 (1-212) x AK065775 (1-3353)

Qy      33 ArgProLeuSerProProProAlaAla 41
           |||||||||||||||||||||||||||
Db     283 CGCCCGCTTCTCCGCCACCGGCAGCC 309

RESULT 97
BC101314/c
LOCUS          BC101314                3699 bp    mRNA            linear   PRI 02-AUG-2005
DEFINITION     Homo sapiens cDNA clone IMAGE:4024855.
ACCESSION      BC101314
VERSION        BC101314.1  GI:71680577
KEYWORDS       .
SOURCE         Homo sapiens (human)
  ORGANISM     Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE      1 (bases 1 to 3699)
  AUTHORS      Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
               Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
               Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
               Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
               Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
               Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
               Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
               Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
               Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
               McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
               Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
               Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
               Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
               Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
               Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
               Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
               Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
               Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  CONSRTM      Mammalian Gene Collection Program Team
  TITLE        Generation and initial analysis of more than 15,000 full-length
               human and mouse cDNA sequences
  JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED      12477932
REFERENCE      2 (bases 1 to 3699)
  CONSRTM      NIH MGC Project
  TITLE        Direct Submission
  JOURNAL      Submitted (01-AUG-2005) National Institutes of Health, Mammalian
               Gene Collection (MGC), Bethesda, MD 20892-2590, USA
  REMARK       NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT        Contact: MGC help desk
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Baylor Human Genome Sequencing Center
               cDNA Library Preparation: Baylor Human Genome Sequencing Center
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Sequencing Group at the Stanford Human Genome

TITLE          Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
               Kojima,K., Namiki,N., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
               Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
               Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
               Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
               Kurosaki,T., Kodama,T., Masuda,H., Yamagata,H., Xie,Q., Lu,M.,
               Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
               Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
               Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
               Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
               Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
               Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
               Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
               Yoshino,M. and Hayashizaki,Y.
TITLE          Collection, mapping, and annotation of over 28,000 cDNA clones from
               japonica rice
JOURNAL        Science 301 (5631), 376-379 (2003)
PUBMED         12869764
REFERENCE      2 (bases 1 to 3353)
AUTHORS        Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
               Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
               Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
               Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
               Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa-Hirozane,T., Itoh,M., Kagawa,I.,
               Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
               Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
               Kodama,T., Kojima,K., Kojima,Y., Kondo,Y., Kondo,S., Konno,H., Kouda,M.,
               Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
               Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
               Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,K.,
               Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K.,
               Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
               Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saito,H., Sakai,C., Sakai,K.,
               Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
               Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,S., Sugano,S.,
               Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
               Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
               Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
               Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
               Yoshimura,A.
TITLE          Direct Submission
JOURNAL        Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
               Agrobiological Sciences, Department of Molecular Genetics, Head of
               Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
               305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
               Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT        This clone is one of the 28K full-length cDNA clones from japonica
               rice.
               URL : http://cdna01.dna.affrc.go.jp/cDNA/
               NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,
               Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
               Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
               Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and
               Yamamoto,M.
               FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y.,
               Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
               Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
               Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S.,
               Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
               Yoshimura,A., Suzuki,Y., Matsubara,K. and Murakami,K.
               Genome Exploration Research Group in Riken Genomic Sciences Center
               and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
               Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
               Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
               Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
               Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
               Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
               Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
               Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,
               Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
               Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
               Sogabe,S., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
               Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,

```
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site:        http://www-shgc.stanford.edu
Contact:         (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAM Plate: 7 Row: m Column: 7.
FEATURES        Location/Qualifiers
     source     1..3699
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4024855"
                /tissue_type="PCR rescued clones"
                /clone_lib="NIH_MGC_283"
                /note="Vector: pCR-Blunt II-TOPO with reversed insert;
                Clone identification sequence tag: NA"
ORIGIN

Alignment Scores:
Pred. No.:              481            Length:         3699
Score:                 9.00           Matches:           9
Percent Similarity:    100.0%         Conservative:      0
Best Local Similarity: 100.0%         Mismatches:        0
Query Match:           4.2%           Indels:            0
DB:                    8              Gaps:              0

US-09-989-890-238 (1-212) x BC101314 (1-3699)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
         |||||||||||||||||||||||||||
Db  131 TTGCCATCCTCCTGGCCTTGTTGTCC 105

RESULT 98
CQ596190
LOCUS       CQ596190                 3840 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 23948 from Patent WO0171042.
ACCESSION   CQ596190
VERSION     CQ596190.1  GI:41652773
KEYWORDS
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 23948 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES        Location/Qualifiers
     source     1..3840
                /organism="Drosophila sp."
                /mol_type="unassigned DNA"
                /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              497            Length:         3840
Score:                 9.00           Matches:           9
Percent Similarity:    100.0%         Conservative:      0
Best Local Similarity: 100.0%         Mismatches:        0
Query Match:           4.2%           Indels:            0
DB:                    6              Gaps:              0

US-09-989-890-238 (1-212) x CQ596190 (1-3840)

Qy   54 LeuProSerAlaSerAlaAlaAlaGly 62
         |||||||||||||||||||||||||||
Db  567 CTACCATCCGCCAAGCGCGCCGCCGGT 593


RESULT 99
MUSERPB
LOCUS       MUSERPB                  3956 bp    DNA     linear   ROD 14-FEB-1996
DEFINITION  Mouse rearranged erythropoietin (epo) gene, 5' flank.
ACCESSION   M26651
VERSION     M26651.1  GI:193088
KEYWORDS    B1 repeat; open reading frame; promoter.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3956)
  AUTHORS   Beru,N., McDonald,J. and Goldwasser,E.
  TITLE     Activation of the erythropoietin gene due to the proximity of an
            expressed gene
  JOURNAL   DNA 8 (4), 253-259 (1989)
   PUBMED   2766927
COMMENT     Original source text: Mus musculus (strain ICFW) (clone: 18C.)
            adult spleen DNA.
            Draft entry and computer-readable copy of sequence [1] kindly
            submitted by N.Beru, 02-AUG-1989.
FEATURES        Location/Qualifiers
     source     1..3956
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /strain="ICFW"
                /db_xref="taxon:10090"
                /clone="18C."
                /cell_line="erythroleukemic cell line IW32"
                /tissue_type="spleen"
                /dev_stage="adult"
     mRNA       complement(<1..570)
                /product="unidentified mRNA"
     repeat_region  complement(251..383)
                /note="B1 repeat element"
     CDS        complement(359..544)
                /pseudo
                /note="ORF; putative"
                /codon_start=1
     protein_bind   1188..1193
                /bound_moiety="Sp1  3"
     protein_bind   1496..1501
                /bound_moiety="Sp1  3"
ORIGIN

Alignment Scores:
Pred. No.:              509            Length:         3956
Score:                 9.00           Matches:           9
Percent Similarity:    100.0%         Conservative:      0
Best Local Similarity: 100.0%         Mismatches:        0
Query Match:           4.2%           Indels:            0
DB:                    9              Gaps:              0

US-09-989-890-238 (1-212) x MUSERPB (1-3956)

Qy   31 LeuSerArgProLeuSerProPro 39
         ||||||||||||||||||||||||
Db 1994 CTGTCCGCCCCCCTTTCCCCACCACCA 2020

RESULT 100
AF225950/c
LOCUS       AF225950                 4109 bp    DNA     linear   PRI 06-DEC-2001
DEFINITION  Homo sapiens HNF-3/forkhead-like protein 1 (HFH1) gene, complete
            cds.
ACCESSION   AF225950
VERSION     AF225950.1  GI:12655883
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
```

```
REFERENCE    Hominidae; Homo.
             1 (bases 1 to 4109)
AUTHORS      Bieller,A., Pasche,B., Frank,S., Glaser,B., Kunz,J., Witt,K. and
             Zoll,B.
TITLE        Isolation and Characterization of the Human Forkhead Gene FOXQ1
JOURNAL      DNA Cell Biol. 20 (9), 555-561 (2001)
PUBMED       11747606
REFERENCE    2 (bases 1 to 4109)
AUTHORS      Bieller,A., Pasche,B. and Zoll,B.
TITLE        Direct Submission
JOURNAL      Submitted (17-JAN-2000) Institute of Human Genetics, University of
             Goettingen, Heinrich-Dueker-Weg 12, Goettingen 37073, Germany
FEATURES             Location/Qualifiers
     source          1..4109
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p23-25"
     gene            <1737..>2948
                     /gene="HFH1"
     mRNA            <1737..>2948
                     /gene="HFH1"
                     /product="HNF-3/forkhead-like protein 1"
     CDS             1737..2948
                     /gene="HFH1"
                     /note="forkhead transcription factor; similar to Mus
                     musculus HNF-3/forkhead homolog 1 like encoded by GenBank
                     Accession Number AF010405 and Rattus norvegicus
                     HNF-3/forkhead homolog-1 encoded by GenBank Accession
                     Number L13201"
                     /codon_start=1
                     /product="HNF-3/forkhead-like protein 1"
                     /protein_id="AAK00639.1"
                     /db_xref="GI:12655884"
                     /translation="MKLEVFVPRAAHGDKQGSDLEGAGGSDAPSPLSAAGDDSLGSDG
                     DCAANSPAAGGARDPPGDGEQSAGGGPGAEEAIPAAAAAVVAEGAEAGAAGPGAGG
                     AGSGEGARSKPYTRRPKPPYSYIALIAMAIRDSAGGRLTLAEINEYLMGKFPFFRGSY
                     TGWRNSVRHNLSLNDCFVKVLRDPSRPWGKDNYWMLNPNSEYTFADGVFRRRRKRLSH
                     RAPVPAPGLRPEEAPGLPAAPPAPAPASPRMRSPARQEERASPAGKFSSSFAIDSI
                     LRKPFRSRRLRDTAPGTTLQWGAAPCPPLPAPPALLPAAPCRALLPLCAYGAGEPARL
                     GAREAEVPPTAPLLLAPLPAAAPAKPLRGPAAGGAHLYCPLRLPAALQAASVRRPGP
                     HLPYPVETLLA"
     misc_feature    2058..2399
                     /gene="HFH1"
                     /note="Region: DNA binding domain"
ORIGIN
```

Alignment Scores:
Pred. No.:   526   Length: 4109
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x AF225950 (1-4109)

```
Qy     59 AlaAlaAlaGlyTyrIleAlaSerSerAla 67
          ||||||||||||||||||||||||||||||
Db   1981 GCTGCTGCCGGGATCGCCTCCTCCGCG 1955
```

Search completed: March 17, 2006, 07:58:07
Job time : 5514 secs

THIS PAGE LEFT BLANK

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

Run on: March 17, 2006, 05:58:11 ; Search time 589 Seconds
(without alignments)
2398.837 Million cell updates/sec

Title: US-09-989-890-238
Perfect score: 212
Sequence: 1 SPHQAAAPVDQTPRTLATMG........RRWAVAPCRABKLMCSSSRS .212

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 496997 seqs, 333234638 residues
Word size: 1
Total number of hits satisfying chosen parameters: 9987775
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/abs/ABSSWEB spool/US09989890/runat_16032006_095513_17502/app_query.fasta_1
-DB=N Geneseq -QFMT=fastap -SUFFIX=oligo_p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=500 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=100 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss05h
-USER=US09989890 @CGN 1 1.727 @runat 16032006_095513 17502 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	212	100.0	791	14	ABA19545	Aea19545 Novel hum
2	212	100.0	1713	8	ABT13390	Abt13390 Breast sp
3	212	100.0	2392	10	ADD01260	Add01260 Human nuc
4	115	54.2	1785	10	ADB62645	Adb62645 Human cDN

Hit list (continued):

No.	c	Count	Score	Length	DB	Accession	Description
5		81	38.2	427	5	AAF66376	Novel hum
6		69	32.5	654	6	ABT07645	Human bre
7		59	27.8	1977	6	ABX92075	Lung spec
8		59	27.8	1977	6	ABX92014	Lung spec
9		11	5.2	4583	4	AAK67070	Human imm
10		11	5.2	261817	11	ACN45182	Human gen
11	c	10	4.7	1635	12	ADQ63387	Novel hum
12	c	10	4.7	2816	13	ADQ64939	Novel hum
13		10	4.7	3421	13	ADR83528	Human for
14		10	4.7	5722	13	ACF87489	Human SIR
15		10	4.7	5722	13	ACF87572	Human SIR
16		10	4.7	5722	13	ABV94143	Breast ca
17		9	4.2	5723	6	ADN06031	Antipsori
18		9	4.2	5723	12	ADN97714	Human fox
19		9	4.2	5723	12	ADN40650	Human for
20		9	4.2	5723	12	ACN39628	Tumour-as
21		9	4.2	5723	13	ADY14881	DNA encod
22		9	4.2	5769	14	ADF81591	Leukaemia
23		9	4.2	5833	10	AAS85053	DNA encod
24	c	9	4.2	239	5	AAS74720	DNA encod
25	c	9	4.2	239	8	ACD05845	Novel hum
26	c	9	4.2	496	2	ACH16725	Human adu
27		9	4.2	546	4	AAX37513	Human sec
28		9	4.2	549	4	ABK43888	DNA encod
29		9	4.2	549	5	AAS34111	Human cDN
30	c	9	4.2	549	5	ADM19666	Novel hum
31		9	4.2	549	12	ADI54275	cDNA enco
32		9	4.2	562	6	ABL89926	Human pol
33		9	4.2	685	12	AAH98991	Rat EST-d
34		9	4.2	717	12	ADJ43042	Plant cDN
35		9	4.2	737	4	ADJ43040	Plant cDN
36		9	4.2	792	4	AAH99445	Human pro
37	c	9	4.2	824	12	ADJ43041	Plant cDN
38		9	4.2	1119	14	ADZ14553	DNA (LOC2
39	c	9	4.2	1209	6	ABQ81557	Gene up-r
40	c	9	4.2	1209	12	ADL70222	Colon can
41	c	9	4.2	1209	12	ADN59614	Colon neo
42	c	9	4.2	1377	14	ACL71856	M. xanthu
43	c	9	4.2	1436	3	AAC52650	Arabidops
44		9	4.2	1553	5	ADM19426	Novel hum
45		9	4.2	1564	5	ABK43589	DNA encod
46		9	4.2	1564	12	ADI53976	cDNA enco
47		9	4.2	1644	8	ABS56295	Human tra
48	c	9	4.2	1716	14	ACL72571	M. xanthu
49		9	4.2	1818	6	AAL44695	Human tra
50		9	4.2	1912	13	ADR86567	UST3-LIKE
51		9	4.2	1960	13	ADN33172	Human tra
52		9	4.2	2071	13	AAS47648	Bacterial
53		9	4.2	2238	13	ADN33202	Human tra
54		9	4.2	2245	6	AAL44696	Human tra
55	c	9	4.2	2256	5	AAS79799	DNA encod
56	c	9	4.2	2275	13	ACN43045	Human dia
57		9	4.2	3840	4	ABL17805	Drosophil
58		9	4.2	5906	4	ABL17804	Drosophil
59		9	4.2	7874	14	ACL64405	M. xanthu
60		9	4.2	7880	4	AAS28527	Genomic s
61		9	4.2	7880	4	AAS28526	Genomic s
62		9	4.2	7880	4	ADG41723	Human res
63		9	4.2	7880	10	ADG41722	Human res
64	c	9	4.2	7880	10	ADI97497	Human res
65	c	9	4.2	9198	11	ADI97496	Human res
66	c	9	4.2	9771	11	ACL64503	M. xanthu
67	c	9	4.2	10432	14	ABX10394	DNA encod
68	c	9	4.2	10432	8	ADI02721	Human cDN
69	c	9	4.2	10432	10	ADJ56543	Murine cD
70	c	9	4.2	10432	10	ADE76855	Human cDN
71	c	9	4.2	10432	12	ADL12292	Human cDN
72	c	9	4.2	10433	12	ADV85846	Human ste
73	c	9	4.2	10433	6	ABS62800	Prostate
74	c	9	4.2	11435	3	AAA64659	cDNA sequ
75	c	9	4.2	11435	14	ADY16694	DNA encod
76	c	9	4.2	11466	12	ADQ22788	Human sof
77	c	9	4.2	12493	3	AAZ43872	Human cel

Left block (rows 78–150)

c	#	Description	ID	num	value	scale	col
c	78	Ade53916 Human pro	ADE53916	10	12494	4.2	9
c	79	Aaa64658 cDNA segu	AAA64658	3	12515	4.2	9
c	80	Abl69122 Kidney ca	ABL69122	6	12515	4.2	9
c	81	Add12702 Human cDN	ADD12702	10	12515	4.2	9
c	82	Adg89365 Cancer de	ADG89365	11	12515	4.2	9
c	83	Adn95435 Human BBC	ADN95435	13	12515	4.2	9
c	84	Adu06058 Novel bro	ADU06058	14	12515	4.2	9
	85	Adx08067 Cyclin-de	ADX08067	14	12515	4.2	9
c	86	Ady16692 DNA encod	ADY16692	14	12515	4.2	9
c	87	Ady16690 DNA encod	ADY16690	14	12515	4.2	9
c	88	Ady61174 Breast ca	ADY61174	14	12515	4.2	9
c	89	Adz09648 Human bre	ADZ09648	14	12515	4.2	9
	90	Aca03928 cDNA down	ACA03928	8	12632	4.2	9
c	91	Acn39127 Tumour-as	ACN39127	13	12674	4.2	9
c	92	Adf81724 Leukaemia	ADF81724	10	12747	4.2	9
c	93	Adf81723 Leukaemia	ADF81723	7	12747	4.2	9
c	94	Adg88601 Human PRO	ADG88601	12	48396	4.2	9
	95	Adx70387 Human Pro	ADX70387	14	49275	4.2	9
	96	Abs56296 Human tra	ABS56296	8	73544	4.2	9
	97	Adc85298 Human Egr	ADC85298	10	96599	4.2	9
c	98	Ada02819 Mouse Sos	ADA02819	9	96600	4.2	9
c	99	Adb72557 Mouse Sos	ADB72557	10	96600	4.2	9
c	100	Adm74414 Murine ca	ADM74414	12	96600	4.2	9
c	101	Abk83569 Human cDN	ABK83569	6	122888	4.2	9
c	102	Adq97284 Mouse can	ADQ97284	12	215248	4.2	9
	103	Abz04549 Human leu	ABZ04549	6	50	3.8	8
	104	Abz00078 Human leu	ABZ00078	6	50	3.8	8
c	105	Adp10107 50-mer ol	ADP10107	12	50	3.8	8
c	106	Aaq49577 Chloropla	AAQ49577	6	55	3.8	8
c	107	Aaq49578 Chloropla	AAQ49578	6	55	3.8	8
c	108	Aaq95018 Chloropla	AAQ95018	2	55	3.8	8
	109	Aaq95017 Chloropla	AAQ95017	2	55	3.8	8
c	110	Aat11019 Corn chlo	AAT11019	8	55	3.8	8
	111	Ada15993 Corn chlo	ADA15993	8	55	3.8	8
c	112	Ada15992 Corn chlo	ADA15992	8	55	3.8	8
c	113	Ach03727 Corn chlo	ACH03727	9	55	3.8	8
	114	Ach03728 Corn chlo	ACH03728	9	55	3.8	8
c	115	Adb23076 Corn chlo	ADB23076	9	55	3.8	8
	116	Adb23077 Corn chlo	ADB23077	9	55	3.8	8
c	117	Adw71719 Corn chlo	ADW71719	14	55	3.8	8
c	118	Adw71720 Corn chlo	ADW71720	14	75	3.8	8
	119	Adj83984 Neisseria	ADJ83984	12	131	3.8	8
c	120	Adv70570 Human DNA	ADV70570	14	143	3.8	8
	121	Adv70571 Human DNA	ADV70571	14	143	3.8	8
	122	Adv70569 Human DNA	ADV70569	14	144	3.8	8
	123	Adj14156 Region of	ADJ14156	11	179	3.8	8
c	124	Ach89925 Human gen	ACH89925	12	180	3.8	8
c	125	Aac08178 Human sec	AAC08178	3	186	3.8	8
c	126	Aai25031 Probe #14	AAI25031	4	186	3.8	8
	127	Aba70629 Human foe	ABA70629	4	186	3.8	8
c	128	Aai50796 Probe #19	AAI50796	4	186	3.8	8
c	129	Aba37193 Probe #15	ABA37193	4	186	3.8	8
c	130	Aak44819 Human bon	AAK44819	4	186	3.8	8
	131	Aak18875 Human bra	AAK18875	4	186	3.8	8
	132	Abs44479 Human liv	ABS44479	6	186	3.8	8
	133	Abs19058 Human gen	ABS19058	11	186	3.8	8
c	134	Adj14171 190bp reg	ADJ14171	14	190	3.8	8
c	135	Adu74098 Liver X r	ADU74098	10	192	3.8	8
c	136	Ady51537 Human liv	ADY51537	13	192	3.8	8
c	137	Ade53990 Human pro	ADE53990	10	201	3.8	8
c	138	Ads36592 Human aut	ADS36592	2	201	3.8	8
c	139	Aat21399 Human gen	AAT21399	12	211	3.8	8
c	140	Ach93942 Human gen	ACH93942	6	220	3.8	8
c	141	Abq98171 Mouse ES	ABQ98171	8	227	3.8	8
c	142	Abx76318 Lung canc	ABX76318	12	270	3.8	8
c	143	Adn04681 Antipsori	ADN04681	12	270	3.8	8
c	144	Aav87697 EST clone	AAV87697	13	270	3.8	8
	145	Aba13191 Human ner	ABA13191	5	274	3.8	8
c	146	Acf57518 Human p16	ACF57518	12	338	3.8	8
c	147	Adj62876 Human p16	ADJ62876	12	340	3.8	8
c	148	Adu81020 Human tum	ADU81020	13	340	3.8	8
c	149	Ady54047 Nucleotid	ADY54047	14	340	3.8	8
c	150	Aea31111 p16 promo	AEA31111	14	340	3.8	8

Right block (rows 151–221)

c	#	Description	ID	num	value	scale	col
	151	Aea19440 Novel hum	AEA19440	14	348	3.8	8
c	152	Abl69313 Prostate	ABL69313	6	349	3.8	8
	153	Aaa82054 N. mening	AAA82054	3	369	3.8	8
	154	Abl66256 Lung canc	ABL66256	6	382	3.8	8
	155	Abl62850 Breast ca	ABL62850	6	382	3.8	8
	156	Abl63074 Breast ca	ABL63074	6	382	3.8	8
	157	Abn96476 Gene #297	ABN96476	9	393	3.8	8
c	158	Ach48117 Human lun	ACH48117	9	395	3.8	8
c	159	Adw81626 MAP3K9 ma	ADW81626	14	411	3.8	8
c	160	Adl40783 Human ova	ADL40783	5	415	3.8	8
	161	Aai82427 Human pol	AAI82427	4	418	3.8	8
	162	Adq50972 Novel can	ADQ50972	13	428	3.8	8
	163	Aba72352 Human foe	ABA72352	4	428	3.8	8
	164	Aai52764 Probe #21	AAI52764	4	428	3.8	8
	165	Aba38168 Probe #16	ABA38168	4	428	3.8	8
	166	Aak46927 Human bon	AAK46927	4	428	3.8	8
	167	Aak20778 Human bra	AAK20778	4	428	3.8	8
	168	Abs46694 Human liv	ABS46694	6	429	3.8	8
	169	Abs21202 Human gen	ABS21202	9	435	3.8	8
	170	Ach48562 Human leu	ACH48562	9	436	3.8	8
c	171	Aca18997 Prokaryot	ACA18997	5	441	3.8	8
c	172	Abv39062 Human pro	ABV39062	5	441	3.8	8
	173	Aaz24597 Human lun	AAZ24597	2	441	3.8	8
	174	Aac65836 Human lun	AAC65836	4	441	3.8	8
	175	Abl49055 Human lun	ABL49055	6	441	3.8	8
	176	Abq92241 Human lun	ABQ92241	6	441	3.8	8
	177	Ada28656 Human lun	ADA28656	6	441	3.8	8
	178	Ades3616 Human lun	ADE53616	10	441	3.8	8
	179	Adh36751 Human lun	ADH36751	10	441	3.8	8
	180	Adm56554 Human lun	ADM56554	12	441	3.8	8
	181	Adn89598 Human lun	ADN89598	12	441	3.8	8
	182	Adu98246 Lung tumo	ADU98246	14	441	3.8	8
	183	Aeb10054 Cancer re	AEB10054	14	442	3.8	8
	184	Aai15812 Probe #57	AAI15812	4	442	3.8	8
	185	Aba58023 Human foe	ABA58023	4	442	3.8	8
c	186	Aai37624 Probe #63	AAI37624	4	442	3.8	8
c	187	Aba27285 Probe #57	ABA27285	4	442	3.8	8
c	188	Aak31748 Human bon	AAK31748	4	442	3.8	8
c	189	Aak06097 Human bra	AAK06097	4	442	3.8	8
c	190	Abs31432 Human liv	ABS31432	4	442	3.8	8
c	191	Abs05504 Human gen	ABS05504	6	445	3.8	8
c	192	Adp66392 Human cDN	ADP66392	12	451	3.8	8
c	193	Ach15449 Human adu	ACH15449	9	463	3.8	8
c	194	Acl55634 Human col	ACL55634	14	464	3.8	8
c	195	Abv14979 Human pro	ABV14979	5	467	3.8	8
c	196	Ach39261 Human foe	ACH39261	9	468	3.8	8
c	197	Aca49485 Prokaryot	ACA49485	8	472	3.8	8
c	198	Aah88733 Plastidic	AAH88733	8	474	3.8	8
c	199	Ach42271 Human foe	ACH42271	9	479	3.8	8
c	200	Aat43202 p16 exon	AAT43202	5	488	3.8	8
c	201	Aba13473 Human ner	ABA13473	2	488	3.8	8
c	202	Ade76340 Human mon	ADE76340	10	488	3.8	8
c	203	Acl63500 Human col	ACL63500	14	488	3.8	8
c	204	Acl63521 Human col	ACL63521	14	492	3.8	8
c	205	Ach76225 Human gen	ACH76225	12	493	3.8	8
c	206	Abv16198 Human pro	ABV16198	5	493	3.8	8
c	207	Ach28173 Human adu	ACH28173	9	509	3.8	8
c	208	Aas58602 cDNA #127	AAS58602	8	512	3.8	8
c	209	Adb82389 Human cDN	ADB82389	12	512	3.8	8
c	210	Adu11294 Solid tum	ADU11294	13	512	3.8	8
c	211	Adl24351 N mening	ADL24351	12	514	3.8	8
c	212	Adl13427 Neisseria	ADL13427	12	525	3.8	8
c	213	Adu79766 Leukaemia	ADU79766	10	525	3.8	8
c	214	Acl35140 Rice stre	ACL35140	11	525	3.8	8
c	215	Abz38961 N. gonorr	ABZ38961	10	526	3.8	8
c	216	Acd05522 cDNA enco	ACD05522	8	527	3.8	8
c	217	Abq14881 Oligonucl	ABQ14881	6	531	3.8	8
	218	Abq14880 Oligonucl	ABQ14880	6	532	3.8	8
c	219	Adf80087 Leukaemia	ADF80087	10	532	3.8	8
c	220	Acc73126 Cat flea	ACC73126	9	543	3.8	8
c	221	Adl09771 Cat flea	ADL09771	12	545	3.8	8

No.	c	Code	n1	n2			Description
224		ACH89654	12	545	3.8	8	Ach89654 Human gen
225		ADN89291	12	549	3.8	8	Adn89291 P16DF DNA
226	c	ACH79329	12	549	3.8	8	Ach79329 Human gen
227	c	AAH97770	5	559	3.8	8	Aah97770 Murine 7-
228		ACH70418	12	563	3.8	8	Ach70418 Human gen
229	c	AAZ40402	12	567	3.8	8	Aaz40402 Human int
230	c	ACL26169	11	567	3.8	8	Acl26169 Rice abio
231		ADF77248	12	568	3.8	8	Adf77248 Interfero
232	c	ACN56532	13	571	3.8	8	Acn56532 Cotton gy
233		ACH69049	12	572	3.8	8	Ach69049 Human gen
234	c	ADU12545	13	576	3.8	8	Adu12545 Solid tum
235	c	ACH78747	12	577	3.8	8	Ach78747 Human gen
236		ABA59811	4	577	3.8	8	Aba59811 Human foe
237		AAI39683	4	577	3.8	8	Aai39683 Probe #83
238		ABA28289	4	577	3.8	8	Aba28289 Probe #67
239		AAK33959	4	577	3.8	8	Aak33959 Human bon
240		AAK08084	4	577	3.8	8	Aak08084 Human bra
241		ABS33764	4	577	3.8	8	Abs33764 Human liv
242		ABS08768	6	577	3.8	8	Abs08768 Human gen
243		AAX83411	2	578	3.8	8	Aax83411 Breast ca
244		AAV68943	2	578	3.8	8	Aav68943 DNA molec
245		AAC80934	3	578	3.8	8	Aac80934 Human bre
246		AAS99780	3	578	3.8	8	Aas99780 Breast tu
247		ABK46824	6	578	3.8	8	Abk46824 Human bre
248		ADA11301	8	578	3.8	8	Ada11301 Human bre
249		ADC15274	11	578	3.8	8	Adc15274 Human bre
250		ACN88200	11	579	3.8	8	Acn88200 Breast ca
251	c	ADD33447	10	585	3.8	8	Add33447 Mouse mit
252		ACH80242	12	589	3.8	8	Ach80242 Human gen
253		AAZ24598	11	600	3.8	8	Aaz24598 Human lun
254		AAC65837	3	600	3.8	8	Aac65837 Human lun
255		ABL49056	6	600	3.8	8	Abl49056 Human lun
256		ABQ92242	6	600	3.8	8	Abq92242 Human lun
257		ADA28657	9	600	3.8	8	Ada28657 Human bre
258		ADE53617	10	600	3.8	8	Ade53617 Human bre
259	c	ADH36752	10	600	3.8	8	Adh36752 Human lun
260	c	ACL28175	11	600	3.8	8	Acl28175 Rice abio
261		ADM56555	12	600	3.8	8	Adm56555 Human lun
262		ADN89599	12	600	3.8	8	Adn89599 Human lun
263		ADU98247	14	600	3.8	8	Adu98247 Lung tumo
264		ABI10055	6	600	3.8	8	Abi10055 Cancer re
265		AAH44236	4	604	3.8	8	Aah44236 Physcomit
266		ABV45996	5	605	3.8	8	Abv45996 Human pro
267		ACN37462	13	617	3.8	8	Acn37462 Tumour-as
268	c	ABX76154	6	619	3.8	8	Abx76154 Lung canc
269	c	ABX10374	8	619	3.8	8	Abx10374 DNA encod
270		ADJ75162	12	619	3.8	8	Adj75162 Marker ge
271		ADN04107	12	619	3.8	8	Adn04107 Antipsori
272		ABQ51293	6	622	3.8	8	Abq51293 Oligonucl
273	c	ABQ51292	6	622	3.8	8	Abq51292 Oligonucl
274	c	ABL66533	6	622	3.8	8	Abl66533 Lung canc
275		ADK70279	12	623	3.8	8	Adk70279 Respirato
276	c	ADM32377	13	623	3.8	8	Adm32377 Human can
277	c	AAH97773	5	633	3.8	8	Aah97773 Murine 7-
278	c	AAC95114	3	642	3.8	8	Aac95114 Cat flea
279	c	ACN45329	13	645	3.8	8	Acn45329 Cotton fl
280	c	AAH97771	5	647	3.8	8	Aah97771 Murine 7-
281	c	AAH97769	5	652	3.8	8	Aah97769 Murine 7-
282		ABA08994	4	653	3.8	8	Aba08994 Human int
283	c	ABK10493	6	660	3.8	8	Abk10493 Human cDN
284		ADN05009	12	660	3.8	8	Adn05009 Antipsori
285		ADZ87558	14	703	3.8	8	Adz87558 DNA methy
286		ADZ77240	13	703	3.8	8	Adz77240 Portion o
287		ABD04770	11	708	3.8	8	Abd04770 Pseudomon
288	c	AAT39040	2	710	3.8	8	Aat39040 Proteinas
289	c	ABS67381	2	710	3.8	8	Abs67381 Neisseria
290	c	ADX60289	13	737	3.8	8	Adx60289 Plant ful
291	c	ADX62605	13	739	3.8	8	Adx62605 Plant ful
292	c	AAH08670	4	740	3.8	8	Aah08670 Human cDN
293		ABQ41088	6	740	3.8	8	Abq41088 Oligonucl
294		ABQ41089	6	740	3.8	8	Abq41089 Oligonucl
295	c	AAC33728	3	753	3.8	8	Aac33728 Arabidops
296	c	ADO82118	13	763	3.8	8	Ado82118 Plant ful
297	c	ADL61988	5	768	3.8	8	Adl61988 Human ova
298	c	AAX39713	2	786	3.8	8	Aax39713 Gastric c
299	c	ADK70277	12	799	3.8	8	Adk70277 Respirato
300	c	ABQ89446	6	806	3.8	8	Abq89446 Human pro
301	c	AAT39042	2	810	3.8	8	Aat39042 Proteinas
302	c	ADX63844	13	815	3.8	8	Adx63844 Plant ful
303	c	AAS62133	6	818	3.8	8	Aas62133 Porcine m
304	c	ABQ56224	6	821	3.8	8	Abq56224 Human ova
305	c	ADO59764	12	829	3.8	8	Ado59764 E. coli m
306	c	AAT39039	2	830	3.8	8	Aat39039 Proteinas
307	c	ACA64711	8	830	3.8	8	Aca64711 N. mening
308	c	ADF43315	10	830	3.8	8	Adf43315 N. mening
309		AEA03009	14	830	3.8	8	Aea03009 Neisseria
310		AAF83374	4	831	3.8	8	Aaf83374 Human SAP
311		ABD15861	11	840	3.8	8	Abd15861 Pseudomon
312		ABQ34587	6	842	3.8	8	Abq34587 Oligonucl
313		ABQ34586	6	842	3.8	8	Abq34586 Oligonucl
314		AAT39041	2	850	3.8	8	Aat39041 Proteinas
315		ADX47013	13	850	3.8	8	Adx47013 Plant ful
316		AAK71514	4	851	3.8	8	Aak71514 Human imm
317		AAK71515	4	851	3.8	8	Aak71515 Human imm
318		ADQ22822	12	869	3.8	8	Adq22822 Human sof
319		AEB64949	14	871	3.8	8	Aeb64949 Rice geno
320		ACL27416	11	885	3.8	8	Acl27416 Rice abio
321		ABK85839	6	894	3.8	8	Abk85839 DNA encod
322	c	ADJ12086	11	894	3.8	8	Adj12086 Maize cDN
323	c	ACL34052	11	894	3.8	8	Acl34052 Rice abio
324	c	ADX32584	13	897	3.8	8	Adx32584 Plant ful
325		ABZ51282	8	909	3.8	8	Abz51282 Aspergill
326		ADD29749	10	909	3.8	8	Add29749 Mouse tum
327		ADS45786	13	912	3.8	8	Ads45786 Bacterial
328		ADX62557	13	916	3.8	8	Adx62557 Plant ful
329		ADX52641	13	926	3.8	8	Adx52641 Plant ful
330	c	AAX39716	2	927	3.8	8	Aax39716 Gastric c
331		ADY85635	14	930	3.8	8	Ady85635 DNA encod
332		AAZ32280	2	934	3.8	8	Aaz32280 Human dun
333		AAA88215	2	934	3.8	8	Aaa88215 pPDE21 hu
334		AAQ47354	2	936	3.8	8	Aaq47354 3' end of
335		ABD17822	11	939	3.8	8	Abd17822 Pseudomon
336		ADX52761	13	945	3.8	8	Adx52761 Plant ful
337		AAH98167	4	957	3.8	8	Aah98167 Rat EST-d
338		ACD05846	8	957	3.8	8	Acd05846 Novel hum
339		ADX10235	13	963	3.8	8	Adx10235 Plant ful
340		AEB66897	14	963	3.8	8	Aeb66897 Rice geno
341		ABQ75304	6	965	3.8	8	Abq75304 Human lun
342		AAA72000	2	987	3.8	8	Aaa72000 Streptomy
343		AAC50604	3	1000	3.8	8	Aac50604 Arabidops
344		AAC36624	3	1007	3.8	8	Aac36624 Arabidops
345		AAH72567	4	1010	3.8	8	Aah72567 Human cer
346		ABZ34915	8	1011	3.8	8	Abz34915 Human gen
347		AAA08593	2	1020	3.8	8	Aaa08593 Human cyt
348		ADQ87400	13	1032	3.8	8	Adq87400 Human tum
349		ADQ85057	13	1032	3.8	8	Adq85057 Human lun
350		ADX63967	13	1032	3.8	8	Adx63967 Plant ful
351		ABL69301	6	1044	3.8	8	Abl69301 Prostate
352		ABL62365	6	1044	3.8	8	Abl62365 Colon ade
353		ABN95743	3	1044	3.8	8	Abn95743 Gene #224
354		ADN95525	11	1044	3.8	8	Adn95525 Human BBC
355		ADN04023	12	1044	3.8	8	Adn04023 Antipsori
356		ACN38888	13	1044	3.8	8	Acn38888 Tumour-as
357		ADS00104	13	1044	3.8	8	Ads00104 Isoform t
358		ADX05863	14	1049	3.8	8	Adx05863 Cyclin-de
359		AAZ48763	3	1049	3.8	8	Aaz48763 Human CID
360		ABL40993	6	1054	3.8	8	Abl40993 Human neu
361		ADO23115	12	1054	3.8	8	Ado23115 Human neu
362		AAF14557	4	1065	3.8	8	Aaf14557 Aspergill
363		ABD17716	11	1065	3.8	8	Abd17716 Pseudomon
364		ADU58598	13	1065	3.8	8	Adu58598 Aspergill
365		ADZ96601	14	1065	3.8	8	Adz96601 Aspergill
366		AAS73988	5	1074	3.8	8	Aas73988 DNA encod
367		AAS74675	5	1083	3.8	8	Aas74675 DNA encod
368		ABN98349	5	1087	3.8	8	Abn98349 Arabidops
369		ACA36091	8	1099	3.8	8	Aca36091 Prokaryot

						Description
c 370	8	3.8	1118	13	ADX61788	Adx61788 Plant ful
c 371	8	3.8	1137	13	ADT47080	Adt47080 Bacterial
c 372	8	3.8	1140	5	AAS74726	Aas74726 DNA encod
c 373	8	3.8	1140	11	ACL34043	Acl34043 Rice abio
c 374	8	3.8	1140	14	AEB56955	Aeb56955 Human pol
c 375	8	3.8	1146	13	AAT15157	Aat15157 CDK4I cyc
c 376	8	3.8	1146	10	AAD64084	Aad64084 Human CDK
c 377	8	3.8	1149	2	AAQ99166	Aaq99166 Human MTS
c 378	8	3.8	1149	2	AAT00737	Aat00737 Multiple
c 379	8	3.8	1149	2	AAV53820	Aav53820 Coding se
c 380	8	3.8	1149	2	AAV11239	Aav11239 Human MTS
c 381	8	3.8	1149	3	AAV70584	Aav70584 Human mul
c 382	8	3.8	1149	3	AAA95634	Aaa95634 Human MTS
c 383	8	3.8	1149	3	AAZ48771	Aaz48771 Human MTS
c 384	8	3.8	1149	3	AAZ39976	Aaz39976 Human mul
c 385	8	3.8	1149	3	AAA39354	Aaa39354 Human MTS
c 386	8	3.8	1149	3	AAA11166	Aaa11166 Human mul
c 387	8	3.8	1149	4	AAF58172	Aaf58172 Multiple
c 388	8	3.8	1149	4	AAS02565	Aas02565 Human mul
c 389	8	3.8	1149	4	AAD04693	Aad04693 Human MTS
c 390	8	3.8	1149	4	AAC83072	Aac83072 5' portio
c 391	8	3.8	1155	2	AAQ14643	Aaq14643 Clone ass
c 392	8	3.8	1155	2	AAT34384	Aat34384 Plasmid p
c 393	8	3.8	1155	3	AAZ32254	Aaz32254 Human tem
c 394	8	3.8	1155	3	AAA88189	Aaa88189 ppDE21 hu
c 395	8	3.8	1158	11	ACL26069	Acl26069 Rice abio
c 396	8	3.8	1161	6	ABK75429	Abk75429 Bacillus
c 397	8	3.8	1163	14	AEB34759	Aeb34759 Genomic D
c 398	8	3.8	1182	6	AAI92636	Aai92636 Human pol
c 399	8	3.8	1187	6	ABZ35437	Abz35437 Human gen
c 400	8	3.8	1190	2	ACN37647	Acn37647 Tumour-as
c 401	8	3.8	1200	6	ABK64802	Abk64802 Human ben
c 402	8	3.8	1200	13	ADX52782	Adx52782 Plant ful
c 403	8	3.8	1209	12	ADI30100	Adi30100 M. methyl
c 404	8	3.8	1216	13	ADR73669	Adr73669 Rice prom
c 405	8	3.8	1218	8	ACC51018	Acc51018 Human bla
c 406	8	3.8	1218	8	ABX76194	Abx76194 Lung canc
c 407	8	3.8	1218	13	ADR73629	Adr73629 Human p16
c 408	8	3.8	1218	11	ACN38874	Acn38874 Tumour-as
c 409	8	3.8	1221	11	ABD10735	Abd10735 Pseudomon
c 410	8	3.8	1228	13	ADX36701	Adx36701 Plant ful
c 411	8	3.8	1235	12	ADQ22927	Adq22927 Human sof
c 412	8	3.8	1251	6	ABK69709	Abk69709 Human lys
c 413	8	3.8	1251	10	ADB67585	Adb67585 Human cDN
c 414	8	3.8	1251	10	ADJ80818	Adj80818 Wilm's tu
c 415	8	3.8	1251	11	ADJ83738	Adj83738 Human lys
c 416	8	3.8	1251	12	ADL57626	Adl57626 Human lys
c 417	8	3.8	1251	12	ADO09181	Ado09181 Human lys
c 418	8	3.8	1280	8	ACL38871	Acl38871 Human stre
c 419	8	3.8	1284	8	ABX63184	Abx63184 Human cDN
c 420	8	3.8	1292	6	ABL14835	Abl14835 cDNA enco
c 421	8	3.8	1298	3	AAZ38628	Aaz38628 Human lys
c 422	8	3.8	1310	10	ADI22488	Adi22488 Rat liver
c 423	8	3.8	1313	13	ADX61997	Adx61997 Plant ful
c 424	8	3.8	1320	9	ADA31793	Ada31793 DNA encod
c 425	8	3.8	1320	13	ADX32173	Adx32173 Plant ful
c 426	8	3.8	1323	13	ADX61135	Adx61135 Plant ful
c 427	8	3.8	1332	11	ABD17357	Abd17357 Pseudomon
c 428	8	3.8	1335	8	ACA54365	Aca54365 Prokaryot
c 429	8	3.8	1337	5	AAH88785	Aah88785 Plastidic
c 430	8	3.8	1338	11	ADS45876	Ads45876 Bacterial
c 431	8	3.8	1341	11	ABD17940	Abd17940 Pseudomon
c 432	8	3.8	1347	11	ACH96431	Ach96431 Klebsiell
c 433	8	3.8	1347	13	ADX62401	Adx62401 Plant ful
c 434	8	3.8	1350	13	ADS56461	Ads56461 Bacterial
c 435	8	3.8	1350	13	ADX64163	Adx64163 Plant ful
c 436	8	3.8	1351	13	ADQ24015	Adq24015 Human sof
c 437	8	3.8	1352	13	ADX36433	Adx36433 Plant ful
c 438	8	3.8	1353	13	ADO82085	Ado82085 Plant ful
c 439	8	3.8	1355	13	ADX62438	Adx62438 Plant ful
c 440	8	3.8	1357	13	ADX64129	Adx64129 Plant ful
c 441	8	3.8	1359	8	ADA71177	Ada71177 Rice gene
c 442	8	3.8	1359	11	ACL26865	Acl26865 Rice abio

						Description
c 443	8	3.8	1364	9	ACC59899	Acc59899 Human REM
c 444	8	3.8	1365	11	ABD10803	Abd10803 Pseudomon
c 445	8	3.8	1368	14	ACL71342	Acl71342 M. xanthu
c 446	8	3.8	1371	13	ADX52969	Adx52969 Plant ful
c 447	8	3.8	1373	13	ADX52876	Adx52876 Plant ful
c 448	8	3.8	1373	13	ADX52791	Adx52791 Plant ful
c 449	8	3.8	1377	2	AAQ89546	Aaq89546 Porcine v
c 450	8	3.8	1380	13	ADS63570	Ads63570 Bacterial
c 451	8	3.8	1380	13	ADS63943	Ads63943 Bacterial
c 452	8	3.8	1380	13	ADS64311	Ads64311 Bacterial
c 453	8	3.8	1384	13	ADX54010	Adx54010 Plant ful
c 454	8	3.8	1394	13	ADX33832	Adx33832 Plant ful
c 455	8	3.8	1397	13	ADX52934	Adx52934 Plant ful
c 456	8	3.8	1401	8	ABX70441	Abx70441 DNA encod
c 457	8	3.8	1401	12	ADL24034	Adl24034 Human NOV
c 458	8	3.8	1408	10	ADJ80236	Adj80236 Novel hum
c 459	8	3.8	1410	13	ADX53122	Adx53122 Plant ful
c 460	8	3.8	1420	2	AAT74051	Aat74051 CDK inhib
c 461	8	3.8	1420	2	AAX26220	Aax26220 Human p27
c 462	8	3.8	1420	3	AAA50523	Aaa50523 DNA encod
c 463	8	3.8	1420	3	AAA90923	Aaa90923 Human W3
c 464	8	3.8	1420	3	AAA50488	Aaa50488 DNA encod
c 465	8	3.8	1439	13	ADX09134	Adx09134 Plant ful
c 466	8	3.8	1440	12	ADM46914	Adm46914 Human ses
c 467	8	3.8	1440	12	ADM46908	Adm46908 Human ses
c 468	8	3.8	1446	13	ADT17429	Adt17429 Plant cDN
c 469	8	3.8	1449	11	ABD16530	Abd16530 Pseudomon
c 470	8	3.8	1452	8	ABX70438	Abx70438 DNA encod
c 471	8	3.8	1452	12	ADL24028	Adl24028 Human NOV
c 472	8	3.8	1459	12	ADL98239	Adl98239 Human bra
c 473	8	3.8	1459	14	ADW00595	Adw00595 Human gen
c 474	8	3.8	1470	2	AAQ46546	Aaq46546 B. subtil
c 475	8	3.8	1470	2	AAT02592	Aat02592 Para-nitr
c 476	8	3.8	1470	2	AAT76591	Aat76591 Para-nitr
c 477	8	3.8	1470	2	AAT76594	Aat76594 Modified
c 478	8	3.8	1470	2	AAT76593	Aat76593 Modified
c 479	8	3.8	1470	2	AAT76592	Aat76592 Modified
c 480	8	3.8	1470	2	AAT76598	Aat76598 Modified
c 481	8	3.8	1470	2	AAT76589	Aat76589 Para-nitr
c 482	8	3.8	1470	2	AAT76596	Aat76596 Para-nitr
c 483	8	3.8	1470	2	AAT76597	Aat76597 Modified
c 484	8	3.8	1470	2	AAT76595	Aat76595 Modified
c 485	8	3.8	1470	2	AAT76590	Aat76590 Modified
c 486	8	3.8	1470	2	AAT76599	Aat76599 Modified
c 487	8	3.8	1470	2	AAX80860	Aax80860 Para-Nitr
c 488	8	3.8	1470	2	AAX80863	Aax80863 Para-Nitr
c 489	8	3.8	1470	2	AAX80853	Aax80853 Para-Nitr
c 490	8	3.8	1470	2	AAX80852	Aax80852 Para-Nitr
c 491	8	3.8	1470	2	AAX80846	Aax80846 Para-Nitr
c 492	8	3.8	1470	2	AAX80849	Aax80849 Para-Nitr
c 493	8	3.8	1470	2	AAX80855	Aax80855 Para-Nitr
c 494	8	3.8	1470	2	AAX80857	Aax80857 Para-Nitr
c 495	8	3.8	1470	2	AAX80851	Aax80851 Para-Nitr
c 496	8	3.8	1470	2	AAX80854	Aax80854 Para-Nitr
c 497	8	3.8	1470	2	AAX80848	Aax80848 Para-Nitr
c 498	8	3.8	1470	2	AAX80847	Aax80847 Para-Nitr
c 499	8	3.8	1470	2	AAX80862	Aax80862 Para-Nitr
c 500	8	3.8	1470	2	AAX80859	Aax80859 Para-Nitr

ALIGNMENTS

RESULT 1
AEA19545
ID   AEA19545 standard; cDNA; 791 BP.
XX
AC   AEA19545;
XX
DT   11-AUG-2005 (first entry)
DE   Novel human polynucleotide SEQ ID NO 239.
XX
KW   vulnerary; CNS-gen.; gene therapy; diagnostic; forensic; mapping;

```
KW   DNA purification; protein purification; osteoarthritis; antiarthritic;
KW   osteopathic; musculoskeletal disease; osteoporosis; endocrine disease;
KW   periodontal disease; antiinflammatory; mouth disease; injury;
KW   peripheral neuropathy; Alzheimers disease; neuroprotective; neotropic;
KW   degeneration parkinsons disease; antiparkinsonian; neurological disease;
KW   cerebrovascular ischemia; cerebroprotective; vasotropic;
KW   cardiovascular disease; autoimmune disease; immunosuppressive;
KW   immune disorder; viral infection; virucide; infection; cancer;
KW   cytostatic; neoplasm; gene; ss.
OS   Homo sapiens.
XX
PN   WO2005049806-A2.
XX
PD   02-JUN-2005.
XX
PF   11-MAR-2004; 2004WO-US007412.
XX
PR   14-MAR-2003; 2003US-00389559.
XX
PA   (NUVE-) NUVELO INC.
XX
PI   Tang TY,   Wang J,    Wang ZW,   Zhang J,    Ren F,     Zhou P,    Ma Y;
PI   Ghosh M,   Xue A,     Asundi V,  Zhao Q,    Wang D,    Goodrich R,  Chen R;
PI   Wehrman T, Weng G,    Boyle B;
XX
DR   WPI; 2005-417730/42.
DR   P-PSDB; AEA20112.
XX
PT   New polynucleotide encoding a polypeptide with biological activity,
PT   useful for treating a disease or disorder, e.g. osteoarthritis, burns,
PT   CNS and peripheral disease, stroke, autoimmune disorders, viral
PT   infection, or cancer.
XX
PS   Claim 1; SEQ ID NO 239; 500pp; English.
XX
CC   The invention describes a new isolated polynucleotide (I) encoding a
CC   polypeptide with biological activity comprising: a nucleotide sequence of
CC   SEQ ID NOS: 1-567 (fully defined); a nucleotide sequence that hybridizes
CC   to the sequence of (i) under stringent hybridization conditions; or a
CC   nucleotide sequence having greater than 99% sequence identity with the
CC   sequence of (i). Also described are: a(n) (expression)vector comprising
CC   (I); a host cell genetically engineered to comprise (I) operatively,
CC   associated with a regulatory sequence that modulates expression of the
CC   polynucleotide in the host cell; an isolated polypeptide comprising a
CC   sequence of SEQ ID NOS: 568-1134 (fully defined), where the polypeptide
CC   is: a polypeptide encoded by (I); or a polypeptide encoded by a
CC   polynucleotide hybridizing under stringent conditions with any one of SEQ
CC   ID NOS: 1-567; a composition comprising the polypeptide of (3) and a
CC   carrier; an antibody directed against the polypeptide of (3); a method
CC   for detecting (I) in a sample; a method for detecting the polypeptide of
CC   (3) in a sample; a method for identifying a compound that binds to the
CC   polypeptide of (3); a method of producing the polypeptide of (3); and a
CC   collection of polynucleotides, where the collection comprising of at
CC   least one of SEQ ID NOS: 1-567. (I) is a polynucleotide comprising any of
CC   the sequences of SEQ ID NOS: 1-567 encoding a polypeptide with biological
CC   activity, which comprises any of the amino acid sequence of SEQ ID NOS:
CC   568-1134. All sequences are fully defined in the specification. The
CC   sequences and methods are useful in diagnostics, forensic, and gene
CC   mapping, in identifying of mutations responsible for genetic disorders or
CC   other traits, in assessing biodiversity, and for producing many other
CC   types of data and products dependent on DNA and amino acid sequences. The
CC   composition and method are useful for treating a disease or disorder,
CC   e.g. osteoporosis, osteoarthritis, periodontal disease, burns, CNS and
CC   peripheral diseases, Alzheimer's disease, Parkinson's disease, stroke,
CC   autoimmune disorders, viral infection, or cancer. This sequence encodes a
CC   novel polypeptide of the invention.
XX
SQ   Sequence 791 BP; 149 A; 278 C; 222 G; 142 T; 0 U; 0 Other;
```

Alignment Scores:		
Pred. No.:	1.03e-180	Length:	791
Score:	212.00	Matches:	212

Percent Similarity:	100.0%		Conservative:	0
Best Local Similarity:	100.0%		Mismatches:	0
Query Match:	100.0%		Indels:	0
DB:	14		Gaps:	0

US-09-989-890-238 (1-212) x ABA19545 (1-791)

```
Qy      1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    103 AGCCCCCACCAGCCGCGCCACCCTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 162

Qy     21 GlnArgAlaLeuProSerSerLeuAlaLeuAlaLeuSerArgProLeuSerProProAla 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    163 CAGAGAGCATTACCTTACCTTCATCTCTGCTCTTGCTGAGCCGCGGCCCTTGAGTCCCCACCTGCT 222

Qy     41 AlaCysSerGlyAlaAspProGlyCysGlyAlaGlyLeuProSerAlaSerAlaAla 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    223 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCCGGCCGCT 282

Qy     61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    283 GCCGGGATTGCCTCCAGCCGCTGTGGAGCCTGTGTGCGGCGGATGCAGCCCCTGCTGTCTA 342

Qy     81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    343 CTGAGGACTCCACTGAGGGACTGTGAAGCCAACACTGGGCCAAGGAGCACAATGGGAGTGC 402

Qy    101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    403 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGCGGCGGGATGGCCAGCGGCTCAAGTCAACCA 462

Qy    121 TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    463 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC 522

Qy    141 ArgGlyTyrProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    523 CGAGGGGCTACCCTCCCCAGCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGATC 582

Qy    161 ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    583 CCCCACCATGCGACACAGCCTGCCCAGCACCTTTGCCAGTCCTGGTCGTGGCCTCCGAGG 642

Qy    181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    643 AGTACTATTCTTTCCATGAGTCGGACCTGGACCTGGCCGGAGATGGGCAGTGGCTCCATGT 702

Qy    201 ArgAlaGluGlyLysLeuMetCysSerSerArgSer 212
          ||||||||||||||||||||||||||||||||||||
Db    703 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 738
```

```
RESULT 2
ABT13390
ID   ABT13390 standard; DNA; 1713 BP.
XX
AC   ABT13390;
XX
DT   30-JAN-2003   (first entry)
XX
DE   Breast specific related polynucleotide SEQ ID No 105.
XX
KW   Cytostatic; BSP-agonist; BSP-antagonist; vaccine; gene therapy; cancer;
KW   metastatic; breast cancer; breast specific; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277232-A2.
XX
PD   03-OCT-2002.
XX
PF   21-NOV-2001; 2001WO-US043815.
XX
PR   22-NOV-2000; 2000US-0252509P.
```

```
Db   644 AGTACTATTCTTTCCATGAGTCGGACCTGGACCTCCCGGAGATGGGCAGTGGCCTCCATGT 703
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   201 ArgAlaGluLysLeuMetCysSerSerArgSer 212
          ||||||||||||||||||||||||||||||||||
Db   704 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAGC 739
```

```
RESULT 3
ADD01260
ID   ADD01260 standard; cDNA; 2392 BP.
XX
AC   ADD01260;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human nucleic acid-associated protein NAAP-41 cDNA SEQ ID NO:98.
XX
KW   human; nucleic acid-associated protein; NAAP; cytostatic;
KW   antiarteriosclerotic; anti-HIV; antiallergic; cerebroprotective;
KW   antiparkinsonian; anticonvulsant; nootropic; neuroprotective;
KW   antiinflammatory; ophthalmological; thyromimetic; antiarthritic;
KW   hepatotropic; antibacterial; virucide; protozoacide; antiparasitic;
KW   fungicide; gene therapy; cell proliferative disease; cancer;
KW   atherosclerosis; hepatitis; neurological disorder; Parkinson's disease;
KW   Alzheimer's disease; stroke; epilepsy; developmental disorder;
KW   renal tubular acidosis; anaemia; glaucoma; hypothyroidism;
KW   autoimmune disorder; inflammatory disorder; AIDS; allergy;
KW   atopic dermatitis; arthritis; infection; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200305219-A2.
XX
PD   03-JUL-2003.
XX
PF   18-DEC-2002; 2002WO-US041115.
XX
PR   19-DEC-2001; 2001US-0343004P.
PR   11-JAN-2002; 2002US-0347633P.
PR   25-JAN-2002; 2002US-0351749P.
PR   22-FEB-2002; 2002US-0359498P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Azimzai Y,  Baughn MR,  Becha SD,  Borowsky ML,  Burford N;
PI   Elliott VS,  Emerling BM,  Forsythe IJ,  Gorvad AE,  Griffin JA;
PI   Kable AE,  Khare R,  Lal PG,  Lee EA,  Lee SY,  Li JX,  Marquis JP;
PI   Ramkumar J,  Richardson TW,  Sprague WW,  Swarnakar A,  Tang YT;
PI   Chawla NK,  Warren BA,  Yue H;
XX
DR   WPI; 2003-559157/52.
DR   P-PSDB; ADD01203.
XX
PT   New human nucleic acid-associated proteins (NAAP), useful for diagnosing,
PT   treating and preventing diseases or conditions associated with the
PT   aberrant NAAP expression, e.g. cancer, hepatitis, AIDS, atherosclerosis,
PT   infections.
XX
PS   Claim 5; SEQ ID NO 98; 405pp; English.
XX
CC   The present invention describes human nucleic acid-associated proteins
CC   designated NAAP-1 to NAAP-57. The human NAAPs have cytostatic,
CC   antiarteriosclerotic, anti-HIV, antiallergic, cerebroprotective,
CC   antiparkinsonian, anticonvulsant, nootropic, neuroprotective,
CC   antiinflammatory, ophthalmological, thyromimetic, antiarthritic,
CC   hepatotropic, antibacterial, virucide, protozoacide, antiparasitic and
CC   fungicide activities, and can be used in gene therapy. The NAAP protein
CC   and polynucleotide sequences can be used in diagnosing, treating and
CC   preventing diseases or conditions associated with the decreased
CC   expression or overexpression of NAAP, such as cell proliferative diseases
CC   (e.g. cancer, atherosclerosis, hepatitis), neurological disorders
CC   (Parkinson's disease, Alzheimer's disease, stroke, epilepsy),
CC   developmental disorders (renal tubular acidosis, anaemia, glaucoma,
```

```
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S,  Macina RA,  Recipon H,  Pluta J,  Sun Y,  Liu C;
XX
DR   WPI; 2003-018927/01.
XX
PT   New isolated nucleic acid molecule, useful for treating breast cancer,
PT   and diagnosing or monitoring the presence of metastases of breast cancer
PT   in a patient.
XX
PS   Claim 1; Page 250-251; 377pp; English.
XX
CC   The invention relates to a novel isolated nucleic acid molecule
CC   comprising: a sequence encoding a sequence comprising 11-1518 amino acids
CC   ; a sequence comprising 190-8144 bp; or a sequence that selectively
CC   hybridises to, or having at least 60% identity with the 11-1518 amino
CC   acid or 190-8144 nt sequence. The polypeptide and the nucleic acid are
CC   useful for treating breast cancer, and diagnosing or monitoring the
CC   presence of metastases of breast cancer in a patient. The polynucleotides
CC   of the invention can be used to treat disorders by gene therapy. This
CC   polynucleotide represents a breast specific related sequence of the
CC   invention
XX
SQ   Sequence 1713 BP; 344 A; 553 C; 468 G; 348 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:             2.1e-180
Score:                 212.00        Length:          1713
Percent Similarity:    100.0%        Matches:          212
Best Local Similarity: 100.0%        Conservative:       0
Query Match:           100.0%        Mismatches:         0
DB:                         8        Indels:             0
                                     Gaps:               0
```

```
US-09-989-890-238 (1-212) x ABT13390 (1-1713)

Qy     1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrAlaAlaThrMetGly 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   104 AGCCCCCACCAGCCGCGCAGCCACCCGTAGACCACCAGACCCCAAGGACCCTGGCCACATGGGC 163

Qy    21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   164 CAGAGAGAGCCATTACCTTCATCTCTGGCTCTCGTCTGCTGAGCCCCTTGAGTCCCCACCTGCT 223

Qy    41 AlaCysSerGlyAlaAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCCGGCGCCTTCTGCTTCTTCCGCCGCT 283

Qy    61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysByLeu 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   284 GCCGGGATTGCCTCCAGCGCTGTGGAGCCCGTGTGTGCGGGGATGCAGCCCCTGCCTGTCTA 343

Qy    81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyLeuProArgSerThrMetGluCys 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   344 CTTGAGGACTTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAAATGGAGTGC 403

Qy   101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGCGGGCGGATGGCCAGCGGCTCAAGTCAACCA 463

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   464 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTATCCCTACC 523

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   524 CGAGGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCGCCTGCAAGGAGCCACTGGCCGGATC 583

Qy   161 ProHisProCysAspThrAlaCysProAlaProLeuProValLeuValAlaProArg 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   584 CCCCACCCATGCGACACAGCCTGCCCAGCACCTTTGCCCAGTAGTCCTCCTGGCTCCGAGG 643

Qy   181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys 200
```

Left column:

```
CC   hypothyroidism), autoimmune/inflammatory disorders (AIDS, allergies,
CC   atopic dermatitis, arthritis) and infections (e.g. bacterial, viral,
CC   parasitic, protozoal, fungal). The present sequence encodes human NAAP-
CC   41, from the present invention.
XX
SQ   Sequence 2392 BP; 471 A; 743 C; 674 G; 504 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                          2.86e-180      Length:         2392
Score:                              212.00         Matches:         212
Percent Similarity:                 100.0%         Conservative:      0
Best Local Similarity:              100.0%         Mismatches:        0
Query Match:                        100.0%         Indels:            0
DB:                                 10             Gaps:              0

US-09-989-890-238 (1-212) x ADD01260 (1-2392)

Qy      1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 AGCCCCCACCAAGCCGCCGCACCGCACCGGTAGACCAGCAGCGGACCCTGGCCACCATGGC 417

Qy     21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 CAGAGAGCATTACCTTCATCTCTGCTCTCGCTGCTGCTCGAGCCCCTTGAGTCCCCACCTGCT 477

Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCCGCT 537

Qy     61 AlaGlyIleAlaSerArgAlaValGluProValCysGlyValAspAlaAlaCysLeu 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    538 GCCGGGATTGCCTCCAGCGCTGTGGAGCCTGTGTGCGGGGATGCAGCCCTGCCTGTCTA 597

Qy     81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    598 CTGAGGACTTCCACTGAGGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC 657

Qy    101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    658 CCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAGCGGCTCAAGTCAACCA 717

Qy    121 TrpAlaAlaAlaSerAlaThrProMetLeuSerLeuLysAlaSerLeuCysIleProThr 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    718 TGGGCAGCAGCTCAGCTACCCCGGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC 777

Qy    141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProIle 160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    778 CGAGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGATC 837

Qy    161 ProHisProCysAspThrAlaCysProAlaProLeuProValLeuValValAlaProArg 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    838 CCCCACCACCATGCGACACAGCCTGCCCAGCACCTTTGCCAGTAGTCCTGTGGCTCCGAGG 897

Qy    181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaAlaProCys 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    898 AGTACTATTCTTTCCATGAGTCGGACCTGACCTGGACCTGCCGGAGATGGGCAGTGGCTCCATGT 957

Qy    201 ArgAlaGluLysLeuMetCysSerSerSerArgSer 212
            |||||||||||||||||||||||||||||||||||
Db    958 CGAGCCGGAGAAATTGATGTGCTCATCTTCAAGAGC 993

RESULT 4
ADB62645
ID   ADB62645 standard; cDNA; 1785 BP.
XX
AC   ADB62645;
XX
DT   04-DEC-2003 (first entry)
XX
DE   Human cDNA encoding clone KIDNB20157100.
XX
KW   Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW   tissue regeneration; cell regeneration; membrane protein;
```

Right column:

```
KW   signal transduction-related protein; transcription-related protein;
KW   osteoporosis; neurological disease; cancer; tumour.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            90. .1286
FT                  /tag= a
FT                  /product= "Clone KIDNB20157100 protein"
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002BP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) REAS ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI   Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI   Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR   WPI; 2003-450961/43.
DR   P-PSDB; ADB64615.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page 1; 222pp; English.
XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein to treat diseases. The
CC   sequence presented is a cDNA of the invention. Note: Some of the sequence
CC   data for this patent is not represented in the printed specification, but
CC   is based on sequence information supplied by the European Patent Office.
XX
SQ   Sequence 1785 BP; 358 A; 567 C; 504 G; 356 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                          2.87e-93       Length:         1785
Score:                              115.00         Matches:         211
Percent Similarity:                 99.1%          Conservative:      0
Best Local Similarity:              99.1%          Mismatches:        1
Query Match:                        54.2%          Indels:            2
DB:                                 10             Gaps:              0

US-09-989-890-238 (1-212) x ADB62645 (1-1785)

Qy      1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

us-09-989-890-238.oligo_p2n.rng

```
Db    193  AGCCCCCACCAAGCCGCCGCCTAGACCAGAGACCCTGCCACCATGGGC  252
Qy     21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuProProAla  40
Db    253  CAGAGAGCATTACCTTCATCTCTGGCTCTTGCTGAGCCGGCCCTTGAGTCCCCACCTGCT  312
Qy     41  AlaCysSerGlyAspProGlyCysGlyGlyAlaGlyLeuProSerAlaSerAlaAla  60
Db    313  GCCTGCTCTGGCGACCCTGGGTGTGGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCCGCT  372
Qy     61  AlaGlyIleAlaSerSerAlaValAlaGluProValCysGlyAspAlaAlaProAlaCysLeu  80
Db    373  GCCGGGATTGCCTCCAGCCGTGTGGCGGCTGTGGGATGCAGCCCCTGCCTGTCTA  432
Qy     81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
Db    433  CTGAGGACTTCACTGAGGGGACTCTGAGCCCAACTGGGCCAAGGAGCACAATGAGTGC  492
Qy    101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGly-SerSerGlnPr  120
Db    493  CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAGCAG-CTCAAGTCAACC  551
Qy    120  oTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProTh  140
Db    552  ATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGACATCCCTGTGTATCCCTAC  611
Qy    140  rArgGlyProProGlnProLeuMetArgThrProArgProAlaAlaArgSerHisTrpProl  160
Db    612  CCGAGGGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCCAAGGAGCCACTGGCCGAT  671
Qy    160  eProHisProCysAspThrAlaCysProAlaProLeuProValLeuValAlaProAr  180
Db    672  CCCCCACCCATGCGACACAGCCTGTGCCCAGCACCTTTGCCAGTAGTCCTGTGGCTCCGAG  731
Qy    180  gSerThrIleLeuSerMetSerArgThrArgThrCysArgArgTrpAlaValAlaProCy  200
Db    732  GAGTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGCAGTGGCTCCATG  791
Qy    200  sArgAlaGluLysLeuMetCysSerSerSerArgSer  212
Db    792  TCGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC  828

RESULT 5
AAF66376
ID    AAF66376 standard; cDNA; 427 BP.
XX
AC    AAF66376;
XX
DT    09-APR-2001  (first entry)
XX
DE    Novel human polynucleotide, SEQ ID NO: 2132.
XX
KW    Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW    breast cancer; lung cancer; cancer detection; ss.
XX
OS    Homo sapiens.
XX
PN    WO200102568-A2.
XX
PD    11-JAN-2001.
XX
PF    30-JUN-2000; 2000WO-US018374.
XX
PR    02-JUL-1999;   99US-0142310P.
PR    02-JUL-1999;   99US-0142311P.
XX
PA    (CHIR ) CHIRON CORP.
PA    (HYSE-) HYSEQ INC.
XX
PI    Williams LT,   Escobedo J,   Innis MA,   Garcia PD,   Klinger J;
PI    Kassam A,   Reinhard C,   Randazzo F,   Kennedy GC,   Pot D,   Lamson G;
PI    Drmanac R,   Crkenjakov R,   Drmanac S,   Dickson M,   Labat I;
PI    Leshkowitz D,   Kita D,   Garcia V,   Jones LW,   Strache-Crain B;
```

XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing cancer, particularly of the colon or
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences.
XX
PS   Claim 9; Page 854; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 427 BP; 69 A; 161 C; 123 G; 73 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:   Length: 427
Score: 2.64e-63   Matches: 81
Percent Similarity: 81.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 38.2%   Gaps: 0
                  5

US-09-989-890-238 (1-212) x AAF66376 (1-427)

```
Qy      1  SerProHisGlnAlaAlaAlaProValAspProThrLeuAlaThrMetGly  20
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183  AGCCCCCACCAAGCCGCCGCCTAGACCAGAGACCCTGCCACCATGGGC  242
Qy     21  GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProAla  40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243  CAGAGAGCATTACCTTCATCTCTGGCTCTTGCTGAGCCGGCCCTTGAGTCCCCACCTGCT  302
Qy     41  AlaCysSerGlyAspProGlyCysGlyGlyAlaGlyLeuProSerAlaSerAlaAla  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303  GCCTGCTCTGGCGACCCTGGGTGTGGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCCGCT  362
Qy     61  AlaGlyIleAlaSerSerAlaValAlaGluProValCysGlyAspAlaAlaProAlaCysLeu  80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363  GCCGGGATTGCCTCCAGCCGTGTGGCGGCTGTGGGNATGCAGCCCCTGCCTGTCTA  422
Qy     81  Leu 81
           |||
Db    423  CTG 425
```

RESULT 6
ABT07645
ID   ABT07645 standard; cDNA; 654 BP.
XX
AC   ABT07645;
XX
DT   14-NOV-2002 (first entry)
XX
DE   Human breast cancer associated coding sequence SEQ ID NO: 124.
XX
KW   Human; breast cancer specific gene; breast specific protein; breast cancer;
KW   gene therapy; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20264611-A1.

XX
PD   22-AUG-2002.
XX
PF   12-FEB-2002; 2002WO-US004197.
XX
PR   13-FEB-2001; 2001US-0268292P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
PI   Sun Y, Liu C;
XX
DR   WPI; 2002-657582/70.
XX
PT   New breast specific nucleic acids and proteins, useful for identifying,
PT   diagnosing, monitoring, staging, imaging, and treating breast cancer and
PT   non-cancerous disease states in breast tissue, and in gene therapy.
XX
PS   Claim 1; Page 248; 367pp; English.
XX
CC   The present invention provides human breast specific coding sequences and
CC   proteins. These can be used in the diagnosis and treatment of breast
CC   cancer and non-cancerous diseases of the breast. The present sequence is
CC   a coding sequence of the invention
XX
SQ   Sequence 654 BP; 118 A; 232 C; 186 G; 118 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.34e-52   Length: 654
Score: 69.00   Matches: 105
Percent Similarity: 98.1%   Conservative: 0
Best Local Similarity: 98.1%   Mismatches: 1
Query Match: 32.5%   Indels: 2
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x ABT07645 (1-654)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
Db   16 AGCCCCCACCAAGCCGCCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 75

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40
Db   76 CAGAGAGCATTACCTTCATCTCTGGCTCTCTGCTCCTCGACCGGCCCCTTGAGTCCCCCACCTGCT 135

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
Db 136 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCCGCT 195

Qy   61 AlaGlyIleAlaSerSerAlaValGluPro-ValCysGlyAlaAlaProAlaCysLeu 80
Db 196 GCCGGGATTGCCTCCAGCGGCGTGGAGGC-CGTGTGCGGGGATGCAGCCCCTGCCTGTCT 254

Qy   80 uLeuArgThrProLeuLysProThrGlyProArgGlySerThrArgMetGluCy 100
Db 255 ACTGGAGGACTCCCCTGAGGGGACCCACTGCTGAAGCCAACTGGTCCAAGGAGGACAATGGAGTG 314

Qy 100 sProProAlaLeuIleVal 106
Db 315 CCCCCAGCCCTGATCGTG 333

RESULT 7
ABX92075
ID   ABX92075 standard; cDNA; 1977 BP.
XX
AC   ABX92075;
XX
DT   08-MAY-2003 (first entry)
XX
DE   Lung specific nucleic acid (LSNA) #117.
XX
KW   Human; gene therapy; vaccine; lung specific antigen; cancer diagnosis;
KW   cancer monitoring; cancer staging; cancer imaging; lung cancer;

KW   non-cancerous diseases of the lung; transgenic animal; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200268633-A2.
XX
PD   06-SEP-2002.
XX
PF   21-NOV-2001; 2001WO-US043612.
XX
PR   22-NOV-2000; 2000US-0252500P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Macina RA, Recipon H, Chen S, Sun Y, Liu C;
XX
DR   WPI; 2002-713376/77.
XX
PT   New isolated human nucleic acid molecule and polypeptide, useful for
PT   identifying, diagnosing, monitoring, staging, imaging and treating lung
PT   cancer and non-cancerous diseases of the lung.
XX
PS   Claim 1; Page 266-267; 389pp; English.
XX
CC   The invention describes an isolated human nucleic acid (I) encoding any
CC   of 120 10-1533 residue amino acid sequences (S1), given in the
CC   specification, comprising any of 164 179-12421 base pair sequences (S2),
CC   given in the specification. The methods and compositions of the present
CC   invention are useful for identifying, diagnosing, monitoring, staging,
CC   imaging and treating lung cancer and non-cancerous diseases of the lung.
CC   They are also used for identifying lung tissue, monitoring and
CC   identifying and/or designing antagonists of the polypeptide of the
CC   invention, gene therapy, production of transgenic animals and production
CC   of engineered lung tissue for treatment and research. This sequence
CC   encodes a lung specific nucleic acid
XX
SQ   Sequence 1977 BP; 393 A; 631 C; 554 G; 399 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   6.21e-43   Length: 1977
Score: 59.00   Matches: 59
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 27.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x ABX92075 (1-1977)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
Db 357 AGCCCCCACCAAGCCGCCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 416

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40
Db 417 CAGAGAGCATTACCTTCATCTCTGGCTCTCTGCTCCTCGAGTCCCCCACCTGCT 476

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla 59
Db 477 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCC 533

RESULT 8
ABX92014
ID   ABX92014 standard; cDNA; 1977 BP.
XX
AC   ABX92014;
XX
DT   08-MAY-2003 (first entry)
XX
DE   Lung specific nucleic acid (LSNA) #56.
XX
KW   Human; gene therapy; vaccine; lung specific antigen; cancer diagnosis;
KW   cancer monitoring; cancer staging; cancer imaging; lung cancer;
KW   non-cancerous diseases of the lung; transgenic animal; gene; ss.

```
XX
OS   Homo sapiens.
XX
PN   WO200268633-A2.
XX
PD   06-SEP-2002.
XX
PF   21-NOV-2001; 2001WO-US043612.
XX
PR   22-NOV-2000; 2000US-0252500P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Macina RA,  Recipon H,   Chen S,   Sun Y,   Liu C;
XX
DR   WPI; 2002-713376/77.
XX
PT   New isolated human nucleic acid molecule and polypeptide, useful for
PT   identifying, diagnosing, monitoring, staging, imaging and treating lung
PT   cancer and non-cancerous diseases of the lung.
XX
PS   Claim 1; Page 205-206; 389pp; English.
XX
CC   The invention describes an isolated human nucleic acid (I) encoding any
CC   of 120 10-1533 residue amino acid sequences (S1), given in the
CC   specification, comprising any of 164 179-12421 base pair sequences (S2),
CC   given in the specification. The methods and compositions of the present
CC   invention are useful for identifying, diagnosing, monitoring, staging,
CC   imaging and treating lung cancer and non-cancerous diseases of the lung.
CC   They are also used for identifying lung tissue, monitoring and
CC   identifying and/or designing antagonists of the polypeptide of the
CC   invention, gene therapy, production of transgenic animals and production
CC   of engineered lung tissue for treatment and research. This sequence
CC   encodes a lung specific nucleic acid
XX
SQ   Sequence 1977 BP; 393 A; 631 C; 554 G; 399 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                 6.21e-43          Length:              1977
Score:                     59.00             Matches:             59
Percent Similarity:        100.0%            Conservative:        0
Best Local Similarity:     100.0%            Mismatches:          0
Query Match:               27.8%             Indels:              0
DB:                        6                 Gaps:                0

US-09-989-890-238 (1-212) x ABX92014 (1-1977)

Qy        1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      357 AGCCCCCACCAACAGCGCCGCCCGCACCCGTAGACCAGGACCCCAAGGACACCCTGGCCACCATGGGC 416

Qy       21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      417 CAGAGAGCATTACCTTCATCTCTGGCTCTTGCTGAGCCGCCCCTTGAGTCCCCCACCTGCT 476

Qy       41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla 59
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      477 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGGTGCCGGGCCTGCCTTCTGCTTCCGCC 533

RESULT 9
AAK67070
ID   AAK67070 standard; DNA; 4583 BP.
XX
AC   AAK67070;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21882.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
```

PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 21882; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 4583 BP; 1269 A; 1249 C; 1011 G; 1054 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 4583
Score: 17.4   Matches: 11
Percent Similarity: 11.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 4   Gaps: 0

US-09-989-890-238 (1-212) x AAK67070 (1-4583)

Qy   24 LeuProSerSerLeuAlaLeuSerArgPro 34

Db   1621 CTTCCATCCAGCCTCGCTCTCCTGAGCCGCCCC 1653

RESULT 10
ACN45182
ID ACN45182 standard; DNA; 261817 BP.
XX
AC ACN45182;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG14925.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200307826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.

PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0246617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

XX
PS   Claim 1; SEQ ID NO 2002; Opp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US20021825586A1, for which no sequence data was published
XX
SQ   Sequence 261817 BP; 71088 A; 59621 C; 57924 G; 73184 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   723   Length: 261817
Score: 11.00   Matches: 11
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.2%   Indels: 0
DB: 11   Gaps: 0

US-09-989-890-238 (1-212) x ACN45182 (1-261817)

Qy   24 LeuProSerSerLeuAlaLeuLeuSerArgPro 34
         |||||||||||||||||||||||||||||||||
Db   241066 CTTCCATCCAGCCTCGCTTCTCCTGAGCCGCCCC 241098

RESULT 11
ADQ63387/c
ID   ADQ63387 standard; cDNA; 1635 BP.
XX
AC   ADQ63387;
XX
DT   07-OCT-2004   (first entry)
XX
DE   Novel human cDNA sequence #548.
XX
KW   ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
KW   cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   EP1440981-A2.
XX
PD   28-JUL-2004.
XX
PF   21-JAN-2004; 2004EP-00001196.
XX
PR   21-JAN-2003; 2003JP-00102206.
PR   09-MAY-2003; 2003JP-00131392.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI   Yamamoto J, Isono Y, Nagai K, Irie R;
XX
DR   WPI; 2004-535376/52.
DR   P-PSDB; ADQ65575.
XX
PT   Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 548; 2449pp; English.
XX

CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotides and polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a nucleotide
CC   sequence of the invention.
XX
SQ   Sequence 1635 BP; 334 A; 466 C; 505 G; 330 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   53.1   Length: 1635
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x ADQ63387 (1-1635)

Qy   26 SerSerLeuAlaLeuLeuSerArgProLeu 35
         ||||||||||||||||||||||||||||||
Db   1332 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 1303

RESULT 12
ADQ64939/c
ID   ADQ64939 standard; cDNA; 2816 BP.
XX
AC   ADQ64939;
XX
DT   07-OCT-2004   (first entry)
XX
DE   Novel human cDNA sequence #2100.
XX
KW   ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
KW   cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   EP1440981-A2.
XX
PD   28-JUL-2004.
XX
PF   21-JAN-2004; 2004EP-00001196.
XX
PR   21-JAN-2003; 2003JP-00102206.
PR   09-MAY-2003; 2003JP-00131392.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI   Yamamoto J, Isono Y, Nagai K, Irie R;
XX
DR   WPI; 2004-535376/52.
DR   P-PSDB; ADQ67127.
XX
PT   Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 2100; 2449pp; English.
XX
CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotides and polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a nucleotide

```
CC   sequence of the invention.
XX
SQ   Sequence 2816 BP; 522 A; 878 C; 818 G; 598 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.:              87.7       Length:        2816
Score:                 10.00       Matches:       10
Percent Similarity:    100.0%      Conservative:  0
Best Local Similarity: 100.0%      Mismatches:    0
Query Match:           4.7%        Indels:        0
DB:                    12          Gaps:          0

US-09-989-890-238 (1-212) x ADQ64939 (1-2816)

Qy    26 SerSerLeuAlaLeuSerArgProLeu 35
         |||||||||||||||||||||||||||
Db   261 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 232

RESULT 13
ADR83528
ID   ADR83528 standard; DNA; 3421 BP.
XX
AC   ADR83528;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Human fork head domain protein DNA, target gene of miRNA.
XX
KW   human; ds; miRNA; microRNA; ontogenesis; cell therapy; cancer;
KW   immune disease; nerve disorder; amyotrophic lateral sclerosis;
KW   Parkinson's disease; Alzheimer's disease; inflammatory disease;
KW   siRNA silencing precursor; antiinflammatory; immunotherapy;
KW   neuroprotective; antiinflammatory; cytostatic; immunosuppressive; nootropic;
KW   fork head domain protein.
XX
OS   Homo sapiens.
XX
PN   WO2004076622-A2.
XX
PD   10-SEP-2004.
XX
PF   10-FEB-2004; 2004WO-JP001433.
XX
PR   10-FEB-2003; 2003US-0445829P.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX
PI   Taira K, Kawasaki H;
XX
DR   WPI; 2004-653393/63.
XX
PT   Modulating expression of a target gene in a cell, for treating cancer, an
PT   immune disease, or a nerve disorder, comprises introducing into the cell
PT   a polynucleotide that forms a duplex region with an mRNA transcribed from
PT   the target gene.
XX
PS   Claim 9; SEQ ID NO 430; 865pp; English.
XX
CC   This invention relates to a novel method for modulating the expression of
CC   a target gene in a cell. Specifically, it refers to the introduction into
CC   a cell of a polynucleotide that forms a duplex region with an mRNA
CC   transcribed from the target gene, where the duplex region comprises a
CC   mammalian miRNA target region i.e. a non-coding microRNA (miRNA) that
CC   regulates mRNA at a post-transcriptional level. The present invention
CC   describes a method for controlling ontogenesis of a mammal, function of a
CC   mammalian cell, differentiation of a mammalian cell or viability of a
CC   mammalian cell in the post-transcriptional phase, which comprises
CC   introducing a plasmid vector comprising a promoter and nucleic acid
CC   molecule expressing an miRNA or siRNA silencing precursor to the miRNA.
CC   Accordingly, it provides a cell therapy method for treating cancer,
CC   immune disease, nerve disorder (e.g. amyotrophic lateral sclerosis,
CC   Parkinson's disease, or Alzheimer's disease) or an inflammatory disease
CC   by introducing into the cell the miRNA, siRNA silencing precursor to the
CC   miRNA or the plasmid vector. As such, they can be developed into
CC   pharmaceutical compositions that exhibit cytostatic, immunosuppressive,
CC   nootropic, neuroprotective and antiinflammatory activities and hence can
CC   be used for immunotherapy. This polynucleotide sequence is a human target
CC   gene whose expression is modulated by miRNAs of the invention.
XX
SQ   Sequence 3421 BP; 852 A; 859 C; 805 G; 905 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.:             105         Length:        3421
Score:                 10.00       Matches:       10
Percent Similarity:    100.0%      Conservative:  0
Best Local Similarity: 100.0%      Mismatches:    0
Query Match:           4.7%        Indels:        0
DB:                    13          Gaps:          0

US-09-989-890-238 (1-212) x ADR83528 (1-3421)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         |||||||||||||||||||||||||||||
Db   172 GCGGGCCTGCCCTCGGCCTGCCGGCT 201

RESULT 14
ACF87489
ID   ACF87489 standard; DNA; 5722 BP.
XX
AC   ACF87489;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Human SIRS/sepsis diagnostic marker DNA fragment 6349.
XX
KW   Systemic inflammatory response syndrome; SIRS; antibacterial;
KW   immunosuppressive; antiinflammatory; diagnosis; sepsis; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004087949-A2.
XX
PD   14-OCT-2004.
XX
PF   31-MAR-2004; 2004WO-EP003419.
XX
PR   02-APR-2003; 2003DE-01015031.
PR   08-AUG-2003; 2003DE-01036511.
PR   02-SEP-2003; 2003DE-01040395.
XX
PA   (SIRS-) SIRS LAB GMBH.
XX
PI   Russwurm S, Reinhart K, Saluz H, Straube E, Zipfel PF, Deigner H;
XX
DR   WPI; 2004-748070/73.
XX
PT   In vitro detection of systemic inflammatory response syndrome and related
PT   conditions, for e.g. monitoring progression, comprises detecting abnormal
PT   expression of disease-related genes.
XX
PS   Disclosure; Page; 75pp; German.
XX
CC   The invention relates to a novel method for in vitro detection of
CC   systemic inflammatory response syndrome (SIRS). The method comprises
CC   detecting abnormal expression of disease-related genes, or their
CC   associated peptides. The method of the invention demonstrates
CC   antibacterial, immunosuppressive and antiinflammatory applications and
CC   may be used for early differential diagnosis, monitoring progression,
CC   assessing risk, assessing the likely response to treatment and for post
CC   mortem diagnosis of systemic inflammatory response syndrome, sepsis and
CC   sepsis-like conditions. The recombinant or synthetic nucleic acid
CC   sequences of the invention, or derived proteins or peptides, may be
CC   useful as calibrants in assays for the specified diseases, for evaluating
CC   activity or toxicity in screening for active agents and/or for
CC   preparation of agents for treatment or prevention of the specified
CC   diseases. The current sequence is that of a human SIRS/sepsis diagnostic
```

Left column

```
CC   marker DNA fragment of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at ftp.wipo.int/pub/published
CC   pct sequences. Furthermore, a number of arbitrary SEQ ID NO.s are
CC   disclosed within the specification, however, these have not been taken
CC   into account during indexing due to inconsistencies in application and
CC   format
XX
SQ   Sequence 5722 BP; 1454 A; 1366 C; 1290 G; 1612 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:             169        Length:              5722
Score:                 10.00      Matches:             10
Percent Similarity:    100.0%     Conservative:        0
Best Local Similarity: 100.0%     Mismatches:          0
Query Match:           4.7%       Indels:              0
DB:                    13         Gaps:                0

US-09-989-890-238 (1-212) x ACF87489 (1-5722)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db    551 GCGGGCCTGCCCCTCGGCCTGCCGCT 580

RESULT 15
ACF87572
ID   ACF87572 standard; DNA; 5722 BP.
XX
AC   ACF87572;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Human SIRS/sepsis diagnostic marker DNA fragment 6432.
XX
KW   Systemic inflammatory response syndrome; SIRS; antibacterial;
KW   immunosuppressive; antiinflammatory; diagnosis; sepsis; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004087949-A2.
XX
PD   14-OCT-2004.
XX
PF   31-MAR-2004; 2004WO-EP003419.
XX
PR   02-APR-2003; 2003DE-01015031.
PR   08-AUG-2003; 2003DE-01036511.
PR   02-SEP-2003; 2003DE-01040395.
XX
PA   (SIRS-) SIRS LAB GMBH.
XX
PI   Russwurm S, Reinhart K, Saluz H, Straube E, Zipfel PF, Deigner H;
XX
DR   WPI; 2004-748070/73.
XX
PT   In vitro detection of systemic inflammatory response syndrome and related
PT   conditions, for e.g. monitoring progression, comprises detecting abnormal
PT   expression of disease-related genes.
XX
PS   Disclosure; Page; 75pp; German.
XX
CC   The invention relates to a novel method for in vitro detection of
CC   systemic inflammatory response syndrome (SIRS). The method comprises
CC   detecting abnormal expression of disease-related genes, or their
CC   associated peptides. The method of the invention demonstrates
CC   antibacterial, immunosuppressive and antiinflammatory applications and
CC   may be used for early differential diagnosis, monitoring progression,
CC   assessing risk, assessing the likely response to treatment and for post
CC   mortem diagnosis of systemic inflammatory response syndrome, sepsis and
CC   sepsis-like conditions. The recombinant or synthetic nucleic acid
CC   sequences of the invention, or derived proteins or peptides, may be
CC   useful as calibrants in assays for the specified diseases, for evaluating
CC   activity or toxicity in screening for active agents and/or for
```

Right column

```
CC   preparation of agents for treatment or prevention of the specified
CC   diseases. The current sequence is that of a human SIRS/sepsis diagnostic
CC   marker DNA fragment of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at ftp.wipo.int/pub/published
CC   pct sequences. Furthermore, a number of arbitrary SEQ ID NO.s are
CC   disclosed within the specification, however, these have not been taken
CC   into account during indexing due to inconsistencies in application and
CC   format
XX
SQ   Sequence 5722 BP; 1454 A; 1366 C; 1290 G; 1612 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:             169        Length:              5722
Score:                 10.00      Matches:             10
Percent Similarity:    100.0%     Conservative:        0
Best Local Similarity: 100.0%     Mismatches:          0
Query Match:           4.7%       Indels:              0
DB:                    13         Gaps:                0

US-09-989-890-238 (1-212) x ACF87572 (1-5722)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db    551 GCGGGCCTGCCCCTCGGCCTGCCGCT 580

RESULT 16
ABV94143
ID   ABV94143 standard; cDNA; 5723 BP.
XX
AC   ABV94143;
XX
DT   08-JAN-2003  (first entry)
XX
DE   Breast carcinoma related nucleotide sequence SEQ ID NO:134.
XX
KW   Human; breast carcinoma; cancer; tumour; cytostatic; anti-tumour; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200246467-A2.
XX
PD   13-JUN-2002.
XX
PF   07-DEC-2001; 2001WO-IB002811.
XX
PR   08-DEC-2000; 2000US-0254090P.
PR   07-DEC-2001; 2001US-00007926.
XX
PA   (IPSO-) IPSOGEN.
XX
PI   Bertucci F, Houlgatte R, Birnbaum D, Nguyen C, Viens P, Fert V;
XX
DR   WPI; 2002-619023/66.
XX
PT   Novel polynucleotide library useful in molecular characterization of a
PT   carcinoma, comprising a pool of polynucleotide sequences or its
PT   subsequences which are either underexpressed or overexpressed in tumor
PT   cells.
XX
PS   Claim 1; Page 198-200; 401pp; English.
XX
CC   The present invention describes a polynucleotide library (I) useful in
CC   the molecular characterisation of a carcinoma, comprising a pool of
CC   polynucleotides or its subsequences which are either underexpressed or
CC   overexpressed in tumour cells, and correspond to any of the
CC   polynucleotide sequences chosen from the 468 sequences given in ABV94010
CC   to ABV94477. Also described: (1) a polynucleotide array (II) useful for
CC   the prognosis or diagnostic of tumour, comprising (I); and (2) detecting
CC   (M1) differentially expressed polynucleotide sequences which are
CC   correlated with a cancer, involves obtaining a polynucleotide sample from
CC   a patient, and reacting the polynucleotide sample obtained with a probe
```

CC   inmobilised on a solid support, where the probe comprises any combination
CC   of the polynucleotide sequences of (I) or its expression products encoded
CC   by polynucleotide sequences of (I), and detecting the reaction product.
CC   (I) have cytostatic activities and can be used as anti-tumour agents. (I)
CC   is useful in molecular characterisation of a carcinoma. (I) and (II) are
CC   useful for the prognosis or diagnostic of tumour, in differentiating a
CC   normal cell from a cancer cell, detecting a hormone sensitive tumour
CC   cell, differentiating a tumour with lymph nodes from a tumour without
CC   lymph nodes, differentiating antracycline-sensitive tumours from
CC   antracycline-insensitive tumours, and classifying good and poor prognosis
CC   primary breast tumours. (I) is useful for large-scale molecular
CC   characterisation of breast cancer that help in prediction, prognosis and
CC   cancer treatment, and for detecting differentially expressed genes that
CC   correlated with a cancer
XX
SQ   Sequence 5723 BP; 1454 A; 1366 C; 1291 G; 1612 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   169   Length: 5723
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x ABV94143 (1-5723)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||
Db   551 GCGGGCCTGCCCTCGGCTGCCGCT 580

RESULT 17
ADN06031
ID   ADN06031 standard; cDNA; 5723 BP.
XX
AC   ADN06031;
XX
DT   01-JUL-2004 (first entry)
XX
DE   Antipsoriatic cDNA sequence #1254.
XX
KW   ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO2004028479-A2.
XX
PD   08-APR-2004.
XX
PI   Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI   Wu TD;
XX
DR   WPI; 2004-305105/28.
DR   P-PSDB; ADN06032.
XX
PA   (GETH) GENENTECH INC.
XX
PT   New PRO nucleic acid or polypeptide, useful for preparing a
PT   pharmaceutical composition for diagnosing or treating psoriasis in a
PT   mammal.
XX
PS   Claim 1; SEQ ID NO 2426; 3069pp; English.
XX
CC   The invention relates to novel polynucleotide and polypeptides for
CC   treating psoriasis or a sequence having at least 80% identity to the
CC   above sequences. The nucleic acid is useful for preparing a composition
CC   for diagnosing or treating psoriasis in a mammal. This sequence
CC   corresponds to one of the polynucleotides of the invention.
XX

SQ   Sequence 5723 BP; 1454 A; 1366 C; 1291 G; 1612 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   169   Length: 5723
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x ADN06031 (1-5723)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||
Db   551 GCGGGCCTGCCCTCGGCTGCCGCT 580

RESULT 18
ADN97714
ID   ADN97714 standard; DNA; 5723 BP.
XX
AC   ADN97714;
XX
DT   01-JUL-2004 (first entry)
XX
DE   Human foxhead box O1A sequence.
XX
KW   ss; cytostatic; antidiabetic; foxhead box O1A inhibitor;
KW   forkhead box O1A; hyperproliferative disorder; cancer; rhabdomyosarcoma;
KW   diabetes; H-ras gene; antisense; gene expression.
XX
OS   Homo sapiens.
XX
PN   WO2004031350-A2.
XX
PD   15-APR-2004.
XX
PF   25-SEP-2003; 2003WO-US030352.
PR   26-SEP-2002; 2002US-00260203.
XX
PA   (AMGE-) AMGEN INC.
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Dobie KW, Bhanot S, Veniant-Ellison M, Lindberg RA, Shutter JR;
XX
DR   WPI; 2004-330164/30.
XX
PT   New compounds, particularly antisense oligonucleotides, targeted to a
PT   nucleic acid molecule encoding forkhead box O1A, useful for treating
PT   cancer, or type 2 diabetes.
XX
PS   Example 15; SEQ ID NO 4; 146pp; English.
XX
CC   The invention relates to a compound 8-80 nucleobases in length targeted
CC   to a nucleic acid molecule encoding forkhead box O1A, where the compound
CC   is at least 70% complementary to a nucleic acid molecule encoding
CC   forkhead box O1A and modulates expression of forkhead box O1A by at least
CC   10%. The compound is useful for treating an animal having a disease or
CC   condition associated with forkhead box O1A, e.g. a hyperproliferative
CC   disorder (cancer, preferably rhabdomyosarcoma), or type 2 diabetes. This
CC   sequence corresponds to the human forkhead box O1A to which the
CC   oligonucleotides of the invention are targeted.
XX
SQ   Sequence 5723 BP; 1454 A; 1366 C; 1291 G; 1612 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   169   Length: 5723
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x ADN97714 (1-5723)

```
Percent Similarity:      100.0%      Conservative:   0
Best Local Similarity:   100.0%      Mismatches:     0
Query Match:             4.7%        Indels:         0
DB:                      12          Gaps:           0
```

```
Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||||
Db     551 GCGGGCCTGCCCCTCGGCCTGCCGCT 580
```

RESULT 19
ADN40650
ID   ADN40650 standard; DNA; 5723 BP.
XX
AC   ADN40650;
XX
DT   12-AUG-2004 (first entry)
XX
DE   Human forkhead box O1A DNA.
XX
KW   Human; forkhead box O1A; gene; ds; antisense oligonucleotide;
KW   phosphorothioate linkage; 2'-O-methoxyethyl sugar moiety;
KW   5-methylcytosine; hyperproliferative disorder; cancer; rhabdomyosarcoma;
KW   type 2 diabetes; cytostatic; antidiabetic.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   386..2353
FT   /*tag= a
FT   /product= "Human forkhead box O1A"
XX
PN   US2004097459-A1.
XX
PD   20-MAY-2004.
XX
PF   25-SEP-2003; 2003US-00671074.
XX
PR   26-SEP-2002; 2002US-00260203.
XX
PA   (DOBI/) DOBIE K W.
PA   (BHAN/) BHANOT S.
PA   (VENI/) VENIANT-ELLISON M.
PA   (LIND/) LINDBERG R A.
PA   (SHUT/) SHUTTER J R.
XX
PI   Dobie KW, Bhanot S, Veniant-Ellison M, Lindberg RA, Shutter JR;
XX
DR   WPI; 2004-389194/36.
DR   P-PSDB; ADN40823.
DR   GENBANK; NM_002015.
XX
PT   New compounds, particularly antisense oligonucleotides, targeted to a
PT   nucleic acid molecule encoding forkhead box O1A, useful for treating
PT   cancer, or type 2 diabetes.
XX
PS   Example 15; SEQ ID NO 4; 80pp; English.
XX
CC   The invention relates to a compound targeted to a nucleic acid molecule
CC   encoding the human forkhead box O1A polypeptide. The compound is an
CC   antisense oligonucleotide that specifically hybridises with the nucleic
CC   acid and inhibits expression of the polypeptide. The antisense
CC   oligonucleotide comprises at least one modified internucleoside linkage
CC   i.e. a phosphorothioate linkage, at least one modified sugar moiety,
CC   preferably a 2'-O-methoxyethyl sugar moiety, or at least one modified
CC   nucleobase comprising a 5-methylcytosine. The antisense compounds are
CC   useful for modulating the expression of the human forkhead box O1A
CC   polypeptide and in preparation of a composition for treating
CC   hyperproliferative disorders, e.g. cancer, preferably rhabdomyosarcoma,
CC   and type 2 diabetes. This sequence represents DNA encoding the human
CC   forkhead O1A polypeptide of the invention.
XX
SQ   Sequence 5723 BP; 1454 A; 1366 C; 1291 G; 1612 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   169   Length: 5723
Score: 10.00   Matches: 10
```

US-09-989-890-238 (1-212) x ADN40650 (1-5723)

```
Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||||
Db     551 GCGGGCCTGCCCCTCGGCCTGCCGCT 580
```

RESULT 20
ACN39628
ID   ACN39628 standard; cDNA; 5723 BP.
XX
AC   ACN39628;
XX
DT   18-NOV-2004 (first entry)
XX
DE   Tumour-associated antigenic target (TAT) cDNA DNA270254, SEQ ID NO:3896.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200403O615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Wu TD, Zhang Z, Zhou Y;
XX
DR   WPI; 2004-347921/32.
DR   P-PSDB; ABM81512.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 1; SEQ ID NO 3896; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT nucleic acid of the invention

```
XX
SQ     Sequence 5723 BP; 1454 A; 1366 C; 1291 G; 1612 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                  169
Score:                      10.00
Percent Similarity:         100.0%
Best Local Similarity:      100.0%
Query Match:                4.7%
DB:                         13

US-09-989-890-238 (1-212) x ACN39628 (1-5723)

Qy       52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
            ||||||||||||||||||||||||||||||
Db      551 GCGGGCCTGCCCTCGGCCTGGCTGCCGCT 580

RESULT 21
ADY14881
ID   ADY14881 standard; DNA; 5723 BP.
XX
AC   ADY14881;
XX
DT   05-MAY-2005 (first entry)
XX
DE   DNA encoding a PRO polypeptide, SEQ ID NO 687.
XX
KW   Antiinflammatory; Immune disorder; Dermatological; Immunosuppressive;
KW   Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;
KW   Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;
KW   Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;
KW   Antiallergic; ds; gene; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO200501962-A2.
XX
PD   24-FEB-2005.
XX
PF   11-AUG-2004; 2004WO-US026249.
PR   11-AUG-2003; 2003US-0493546P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Abbas A, Clark H, Ouyang W, Williams.MP, Wood WI, Wu TD;
XX
DR   WPI; 2005-182330/19.
XX
PT   New nucleic acid encoding PRO polypeptide, useful for diagnosing and
PT   treating an immune related disorder, e.g. systemic lupus erythematosus,
PT   rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.
XX
PS   Claim 1; SEQ ID NO 687; 158pp; English.
XX
CC   The invention relates to an isolated nucleic acid encoding a PRO
CC   polypeptide. The polypeptide, agonist or an antagonist, antibody,
CC   composition, and method are useful for diagnosing and treating an immune
CC   related disorder, e.g. systemic lupus erythematosus, rheumatoid
CC   arthritis. The present sequence represents a DNA encoding a PRO
CC   polypeptide.
XX
SQ   Sequence 5723 BP; 1454 A; 1366 C; 1291 G; 1612 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                  169            Length:           5723
Score:                      10.00          Matches:          10
Percent Similarity:         100.0%         Conservative:     0
Best Local Similarity:      100.0%         Mismatches:       0
Query Match:                4.7%           Indels:           0
DB:                         14            Gaps:             0

US-09-989-890-238 (1-212) x ADY14881 (1-5723)
```

```
Qy       52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
            ||||||||||||||||||||||||||||||
Db      551 GCGGGCCTGCCCTCGGCCTGGCTGCCGCT 580

RESULT 22
ADF81591
ID   ADF81591 standard; DNA; 5769 BP.
XX
AC   ADF81591;
XX
DT   26-FEB-2004 (first entry)
XX
DE   Leukaemia-related DNA sequence #2147.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
XX
OS   Unidentified.
XX
PN   WO2003039443-A2.
XX
PD   15-MAY-2003.
XX
PF   04-NOV-2002; 2002WO-EP012303.
XX
PR   05-NOV-2001; 2001EP-00126244.
PR   30-APR-2002; 2002EP-00009758.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA   (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA   (HAFE/) HAFERLACH T.
PA   (SCHO/) SCHOCH C.
PA   (KERN/) KERN W.
XX
PI   Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI   Eils R, Brors B, Mergenthaler S;
XX
DR   WPI; 2003-505037/47.
XX
PT   Determining the subtype of leukemia cells and whether a patient sample
PT   contains leukaemia cells or other cells, useful for treating leukemia,
PT   comprises determining the expression profile of a group of markers in a
PT   patient sample.
XX
PS   Disclosure; SEQ ID NO 2147; 2938pp; English.
XX
CC   The present invention relates to a method (M1) for determining the
CC   subtype of leukaemia cells and whether a patient sample contains
CC   leukaemia cells. The method comprises determining the expression profile
CC   of a group of markers in a patient sample. The method is useful for
CC   determining the presence of leukaemia cells, its types or subtypes, and
CC   for the preparation of a medicament for treating leukaemia.
XX
SQ   Sequence 5769 BP; 1485 A; 1367 C; 1291 G; 1613 T; 0 U; 13 Other;

Alignment Scores:
Pred. No.:                  170            Length:           5769
Score:                      10.00          Matches:          10
Percent Similarity:         100.0%         Conservative:     0
Best Local Similarity:      100.0%         Mismatches:       0
Query Match:                4.7%           Indels:           0
DB:                         10            Gaps:             0

US-09-989-890-238 (1-212) x ADF81591 (1-5769)

Qy       52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
            ||||||||||||||||||||||||||||||
Db      551 GCGGGCCTGCCCTCGGCCTGGCTGCCGCT 580

RESULT 23
AAS85053
ID   AAS85053 standard; cDNA; 5833 BP.
XX
```

```
AC   AAS85053;
XX
DT   13-FEB-2002   (first entry)
XX
DE   DNA encoding novel human diagnostic protein #20857.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG20866.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 20857; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have application in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5833 BP; 1482 A; 1383 C; 1331 G; 1637 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                 172           Length:                5833
Score:                     10.00         Matches:               10
Percent Similarity:        100.0%        Conservative:          0
Best Local Similarity:     100.0%        Mismatches:            0
Query Match:               4.7%          Indels:                0
DB:                        5             Gaps:                  0

US-09-989-890-238 (1-212) x AAS85053 (1-5833)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAlaAla 61
         |||||||||||||||||||||||||||||||||
Db   552 GCGGGCCTGCCCTCGGCCTCGGCTGCCGCT 581
```

```
AAS74720/c
ID   AAS74720 standard; cDNA; 239 BP.
XX
AC   AAS74720;
XX
DT   13-FEB-2002   (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10524.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10533.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 10524; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 239 BP; 39 A; 91 C; 59 G; 50 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                 71.4          Length:                239
Score:                     9.00          Matches:               9
Percent Similarity:        100.0%        Conservative:          0
Best Local Similarity:     100.0%        Mismatches:            0
Query Match:               4.2%          Indels:                0
DB:                        5             Gaps:                  0

US-09-989-890-238 (1-212) x AAS74720 (1-239)

Qy    54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
         ||||||||||||||||||||||||||||||
```

. RESULT 24

Db   131 CTTCCTTCAGGGTCAGCAGCAGCTGA 105

RESULT 25
ACD05845/c
ID   ACD05845 standard; cDNA; 239 BP.
XX
AC   ACD05845;
XX
DT   06-AUG-2003 (first entry)
XX
DE   Novel human contig #19.
XX
KW   Human; angiogenesis; cytokine; cell proliferation; pluripotent;
KW   cell differentiation; totipotent; stem cell; transplantation; bio-sensor;
KW   neuroepithelial cell; autoimmune disease; neural cell; genetic disorder;
KW   nerve; brain tissue; central nervous system disease;
KW   peripheral nervous system disease; neuropathy; haematopoiesis; bone;
KW   myeloid disorder; lymphoid cell disorder; platelet disorder; tendon;
KW   regeneration; cartilage; tendon; ligament; nerve tissue growth;
KW   tissue repair; wound healing; burn; ulcer; osteoporosis; cancer;
KW   osteoarthritis; bone degenerative disorder; periodontal disease;
KW   gut protection; lung fibrosis; liver fibrosis; reperfusion injury;
KW   immune deficiency; infection; autoimmune disorder; allergic reaction;
KW   thrombolysis; thrombosis; coagulation disorder; hereditary disorder;
KW   biorhythm; circadian cycle; fertility; metabolism; catabolism; anabolism;
KW   nootropic; neuroprotective; antiparkinsonian; anticonvulsant;
KW   haemostatic; vulnerary; antiulcer; osteopathic; antiarthritic;
KW   vasotropic; immunostimulant; antibacterial; fungicide; immunosuppressive;
KW   antirheumatic; antidiabetic; antiasthmatic; cytostatic; virucide;
KW   expressed sequence tag; EST; SS.
XX
OS   Homo sapiens.
XX
PN   WO200323013-A2.
XX
PD   20-MAR-2003.
XX
PF   13-SEP-2002; 2002WO-US029001.
XX
PR   13-SEP-2001; 200US-0322511P.
PR   12-SEP-2002; 2002US-00243552.
XX
PA   (HYSB-) HYSEQ INC.
XX
PI   Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
XX
DR   WPI; 2003-313249/30.
DR   P-PSDB; AB000768.
XX
PT   Novel nucleic acids and polypeptides for diagnosis, treatment of central
PT   and peripheral nervous system diseases and neuropathies, such as
PT   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
PT   lateral sclerosis.
XX
PS   Example 2; SEQ ID NO 691; 300pp; English.
XX
CC   The present invention relates to the isolation of novel human
CC   polynucleotide sequences and their encoding polypeptides. The novel
CC   polypeptides exhibit activities relating to angiogenesis, cytokine, cell
CC   proliferation, cell differentiation, antiinflammatory, and stem cell
CC   growth factor activities. The polypeptides are involved in the
CC   proliferation, differentiation and survival of pluripotent and totipotent
CC   stem cells, and are useful for re-engineering diseased or diseased
CC   tissues, transplantation, manufacture of bio-pharmaceuticals and
CC   development of bio-sensors. The polypeptides can be used to manipulate
CC   stem cells in culture to give rise to neuroepithelial cells that can be
CC   used to augment or replace cells damaged by illness, autoimmune disease,
CC   accidental damage or genetic disorders. The polypeptides induce the
CC   proliferation of neural cells and regeneration of nerve and brain tissue
CC   and are useful for the treatment of central and peripheral nervous system
CC   diseases and neuropathies, such as Alzheimer's, Parkinson's disease,
CC   Huntington's disease, amyotrophic lateral sclerosis (ALS). The
CC   polypeptides are also involved in chemotactic or chemokinetic activity,
CC   regulation of haematopoiesis and are useful for treating myeloid or
CC   lymphoid cell disorders, platelet disorders such as thrombocytopaenia and
CC   for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC   growth, in tissue repair, healing of burns, incisions, ulcers, for
CC   treating osteoporosis, osteoarthritis, bone degenerative disorders, and
CC   periodontal disease. The polypeptides are also useful for gut protection
CC   or regeneration and treatment of lung or liver fibrosis, reperfusion
CC   injury in various tissues, various immune deficiencies and disorders
CC   including severe combined immunodeficiency (SCID), bacterial or fungal
CC   infections, autoimmune disorders (e.g. multiple sclerosis, rheumatoid
CC   arthritis, diabetes mellitus, myasthenia gravis), allergic reactions and
CC   conditions, such as asthma or other respiratory problems. The
CC   polypeptides are involved in thrombolysis or thrombosis and are useful in
CC   treatment of various coagulation disorders (including hereditary
CC   disorders such as haemophilia) or to enhance coagulation and other
CC   haemostatic events in treating wounds resulting from trauma, surgery or
CC   other causes. The polypeptides exhibit immune stimulating or immune
CC   suppressing activity, and are useful for treating autoimmune diseases or
CC   cancer. They also inhibit the growth, infection or function of infectious
CC   agents such as bacteria, fungi, viruses, effect biorhythms or circadian
CC   cycles of rhythms, fertility of male or female subjects, metabolism,
CC   catabolism, and anabolism. ACD05827-ACD06027 represent novel contigs
CC   assembled using expressed sequence tag (EST) sequences as seeds. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 239 BP; 39 A; 91 C; 59 G; 50 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 239
Score: 71.4   Matches: 9
Percent Similarity: 9.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 4.2%   Gaps: 0

US-09-989-890-238 (1-212) x ACD05845 (1-239)

Qy   54 LeuProSerAlaSerAlaAlaAlaGly 62
          ||||||||||||||||||||||||||||
Db   131 CTTCCTTCAGGGTCAGCAGCAGCTGA 105

RESULT 26
ACH16725/c
ID   ACH16725 standard; cDNA; 496 BP.
XX
AC   ACH16725;
XX
DT   13-OCT-2003 (first entry)
XX
DE   Human adult heart cDNA #1039.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 3937; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 496 BP; 108 A; 157 C; 121 G; 108 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:   140   Length: 496
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x ACH16725 (1-496)

Qy   25 ProSerSerLeuAlaLeuSerArg 33
             ||||||||||||||||||||||||
Db   159 CCGAGCTCCCTGGCTCTACTATCCAGG 133

RESULT 27
AAX37513/c
ID   AAX37513 standard; cDNA; 546 BP.
XX
AC   AAX37513;
XX
DT   06-JUL-1999 (first entry)
XX
DE   Human secreted protein cDNA fragment containing gene 63.
XX
KW   Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW   gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW   immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW   renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW   cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW   pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW   arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO9918208-A1.
XX
PD   15-APR-1999.
XX
PF   01-OCT-1998; 98WO-US020775.
XX
PR   02-OCT-1997; 97US-0060833P.
PR   02-OCT-1997; 97US-0060836P.

PR   02-OCT-1997; 97US-0060837P.
PR   02-OCT-1997; 97US-0060838P.
PR   02-OCT-1997; 97US-0060839P.
PR   02-OCT-1997; 97US-0060843P.
PR   02-OCT-1997; 97US-0060862P.
PR   02-OCT-1997; 97US-0060866P.
PR   02-OCT-1997; 97US-0060874P.
PR   02-OCT-1997; 97US-0060880P.
PR   02-OCT-1997; 97US-0060884P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
PI   Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;
PI   Lafleur DW, Shi Y;
XX
DR   WPI; 1999-264022/22.
DR   P-PSDB; AAY07914.
XX
PT   New isolated human genes and the secreted polypeptides they encode.
XX
PS   Claim 1a; Page 259; 368pp; English.
XX
CC   This invention describes novel isolated human genes and the secreted
CC   proteins they encode. The products of the invention are useful for
CC   preventing, treating or ameliorating medical conditions, e.g. by protein
CC   or gene therapy. Also pathological conditions can be diagnosed by
CC   determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 101 polynucleotides, based on
CC   which tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, tumours,
CC   neurodegenerative disorders, developmental abnormalities and fetal
CC   deficiencies, blood disorders, leukemias, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC   allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC   disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC   transplant rejection, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC   infections and AIDS. The human secreted proteins of the invention are
CC   represented in AAY07852-Y07993 and the encoding nucleic acids are
CC   represented in AAX37451-X37552
XX
SQ   Sequence 546 BP; 193 A; 97 C; 120 G; 136 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   153   Length: 546
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x AAX37513 (1-546)

Qy   127 ThrProMetLeuSerLysAlaSer 135
             ||||||||||||||||||||||||
Db   276 ACACCTATGCTCTCTCCAAAGCTTCA 250

RESULT 28
ABK43888
ID   ABK43888 standard; cDNA; 549 BP.
XX
AC   ABK43888;
XX
DT   05-JUN-2002 (first entry)
XX
DE   DNA encoding novel central nervous system protein #468.
XX
KW   Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW   cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;

```
KW   acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW   adenocarcinoma; reproductive system disorder; testicular feminisation;
KW   endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW   respiratory disorder; renal disorder; kidney failure; blood disorder;
KW   myocardial infarction; wound healing; cell proliferation; skin aging;
KW   food additive; food preservative; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20015318-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001332.
XX
PR   31-JAN-2001; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0271182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
```

```
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0255484P.
PR   27-SEP-2000; 2000US-0255834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
```

PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-581633/65.
DR   P-PSDB; AAU87558.
XX
PT   New isolated nucleic acid encoding a protein for diagnosing, preventing,
PT   treating or ameliorating medical conditions and used as food additives or
PT   preservatives.
XX
PS   Claim 1; SEQ ID NO 478; 837pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule (I) encoding a
CC   novel central nervous system protein. (I) and polypeptides (III) encoded
CC   by (I), are used to treat a medical conditions and in diagnosis of a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC   amylotrophic lateral sclerosis, infections caused by bacteria, viruses
CC   e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC   e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC   adenocarcinomas and irritable bowel syndrome, reproductive system
CC   disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC   and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC   leukaemia, disorders involving neovascularisation e.g. malignancies,
CC   respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC   acute kidney failure and blood related disorders e.g. myocardial
CC   infarction. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:
Pred. No.:   154   Length: 549
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 4   Gaps: 0

US-09-989-890-238 (1-212) x ABK43888 (1-549)

Qy   24 LeuProSerLeuAlaLeuLeuSer 32
         ||||||||||||||||||||||||
Db   349 TTGCCATCCTCTCTGGCCCTGTTGTCTCC 375

RESULT 29
AAS34111
ID   AAS34111 standard; cDNA; 549 BP.
XX
AC   AAS34111;
XX
DT   17-DEC-2001 . (first entry)
XX
DE   Human cDNA encoding a novel foetal antigen, SEQ ID No 635.
XX
KW   Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
KW   immunomodulator; cardiovascular; cytostatic; nephrothropic;
KW   cardiovascular; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; breast neoplasm; cancer;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;

KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200155312-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001321.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0230063P.
PR   14-SEP-2000; 2000US-0230064P.
PR   14-SEP-2000; 2000US-0230065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.

PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.

PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-488782/53.
DR   P-PSDB; AAU21291.
XX
PT   New polynucleotides and polypeptides for diagnosing, treating, preventing
PT   or prognosing e.g. diseases or disorders of the nervous, musculoskeletal,
PT   excretory, gastrointestinal, reproductive, and respiratory systems.
XX
PS   Claim 1; SEQ ID NO 635; 642pp; English.
XX
CC   The invention relates to novel nucleic acids encoding novel human foetal
CC   antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC   by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. The antibodies to the antigens can also be used
CC   in alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC   The polypeptides can also be used to aid wound healing and epithelial
CC   cell proliferation, to prevent skin aging due to sunburn, to maintain
CC   organs before transplantation, for supporting cell culture of primary
CC   tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC   also be used as a food additive or preservative to increase or decrease
CC   storage capabilities, fat content, lipid, protein, carbohydrate,
CC   vitamins, minerals, cofactors and other nutritional components. Numerous
CC   examples of diseases and disorders treated by the nucleic acids and
CC   proteins are given in the specification. The present sequence encodes a
CC   foetal antigen of the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in

Alignment Scores:
Pred. No.:   154   Length: 549
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 5   Gaps: 0

US-09-989-890-238 (1-212) x AAS34111 (1-549)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
              ||||||||||||||||||||||||||||
Db   349 TTGCCATCCTCTCTGGCCTTGTTGTCC 375

RESULT 30
ADM19666
ID   ADM19666 standard; cDNA; 549 BP.
XX
AC   ADM19666;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Novel human channel/transporter gene #223 clone 2.
XX
KW   ds; gene; immunosuppressive; antiarthritic; antirheumatic;
KW   antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW   nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW   opthalmological; gene therapy; channel/transporter protein;
KW   rheumatoid arthritis; neoplasm; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW   epithelial cell proliferation; skin aging; sunburn; transplantation;

```
KW   chemotaxis; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200154472-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001;  2001WO-US001307.
XX
PR   31-JAN-2000;  2000US-0179065P.         PR   25-SEP-2000;  2000US-0234998P.
PR   04-FEB-2000;  2000US-0180628P.         PR   26-SEP-2000;  2000US-0235484P.
PR   24-FEB-2000;  2000US-0184664P.         PR   27-SEP-2000;  2000US-0235834P.
PR   02-MAR-2000;  2000US-0186350P.         PR   27-SEP-2000;  2000US-0235836P.
PR   16-MAR-2000;  2000US-0189874P.         PR   29-SEP-2000;  2000US-0236327P.
PR   17-MAR-2000;  2000US-0190076P.         PR   29-SEP-2000;  2000US-0236367P.
PR   18-APR-2000;  2000US-0198123P.         PR   29-SEP-2000;  2000US-0236368P.
PR   19-MAY-2000;  2000US-0205515P.         PR   29-SEP-2000;  2000US-0236369P.
PR   07-JUN-2000;  2000US-0209467P.         PR   29-SEP-2000;  2000US-0236370P.
PR   28-JUN-2000;  2000US-0214886P.         PR   02-OCT-2000;  2000US-0236802P.
PR   30-JUN-2000;  2000US-0215135P.         PR   02-OCT-2000;  2000US-0237037P.
PR   07-JUL-2000;  2000US-0216647P.         PR   02-OCT-2000;  2000US-0237038P.
PR   07-JUL-2000;  2000US-0216880P.         PR   02-OCT-2000;  2000US-0237039P.
PR   11-JUL-2000;  2000US-0217487P.         PR   02-OCT-2000;  2000US-0237040P.
PR   11-JUL-2000;  2000US-0217496P.         PR   13-OCT-2000;  2000US-0239935P.
PR   14-JUL-2000;  2000US-0218290P.         PR   13-OCT-2000;  2000US-0239937P.
PR   26-JUL-2000;  2000US-0220963P.         PR   20-OCT-2000;  2000US-0240960P.
PR   26-JUL-2000;  2000US-0220964P.         PR   20-OCT-2000;  2000US-0241221P.
PR   14-AUG-2000;  2000US-0224518P.         PR   20-OCT-2000;  2000US-0241785P.
PR   14-AUG-2000;  2000US-0224519P.         PR   20-OCT-2000;  2000US-0241786P.
PR   14-AUG-2000;  2000US-0225213P.         PR   20-OCT-2000;  2000US-0241787P.
PR   14-AUG-2000;  2000US-0225214P.         PR   20-OCT-2000;  2000US-0241808P.
PR   14-AUG-2000;  2000US-0252266P.         PR   20-OCT-2000;  2000US-0241809P.
PR   14-AUG-2000;  2000US-0252267P.         PR   20-OCT-2000;  2000US-0241826P.
PR   14-AUG-2000;  2000US-0252268P.         PR   01-NOV-2000;  2000US-0244617P.
PR   14-AUG-2000;  2000US-0252270P.         PR   08-NOV-2000;  2000US-0246474P.
PR   14-AUG-2000;  2000US-0254447P.         PR   08-NOV-2000;  2000US-0246475P.
PR   14-AUG-2000;  2000US-0257757P.         PR   08-NOV-2000;  2000US-0246476P.
PR   14-AUG-2000;  2000US-0257758P.         PR   08-NOV-2000;  2000US-0246477P.
PR   14-AUG-2000;  2000US-0257759P.         PR   08-NOV-2000;  2000US-0246478P.
PR   18-AUG-2000;  2000US-0262279P.         PR   08-NOV-2000;  2000US-0246523P.
PR   22-AUG-2000;  2000US-0226681P.         PR   08-NOV-2000;  2000US-0246524P.
PR   22-AUG-2000;  2000US-0226868P.         PR   08-NOV-2000;  2000US-0246525P.
PR   22-AUG-2000;  2000US-0271182P.         PR   08-NOV-2000;  2000US-0246526P.
PR   23-AUG-2000;  2000US-0227009P.         PR   08-NOV-2000;  2000US-0246527P.
PR   30-AUG-2000;  2000US-0228924P.         PR   08-NOV-2000;  2000US-0246528P.
PR   01-SEP-2000;  2000US-0292877P.         PR   08-NOV-2000;  2000US-0246532P.
PR   01-SEP-2000;  2000US-0229343P.         PR   08-NOV-2000;  2000US-0246609P.
PR   01-SEP-2000;  2000US-0229344P.         PR   08-NOV-2000;  2000US-0246610P.
PR   01-SEP-2000;  2000US-0229345P.         PR   08-NOV-2000;  2000US-0246611P.
PR   05-SEP-2000;  2000US-0229509P.         PR   08-NOV-2000;  2000US-0246613P.
PR   05-SEP-2000;  2000US-0229513P.         PR   17-NOV-2000;  2000US-0249207P.
PR   06-SEP-2000;  2000US-0230437P.         PR   17-NOV-2000;  2000US-0249208P.
PR   06-SEP-2000;  2000US-0230438P.         PR   17-NOV-2000;  2000US-0249209P.
PR   08-SEP-2000;  2000US-0231242P.         PR   17-NOV-2000;  2000US-0249210P.
PR   08-SEP-2000;  2000US-0231243P.         PR   17-NOV-2000;  2000US-0249211P.
PR   08-SEP-2000;  2000US-0231244P.         PR   17-NOV-2000;  2000US-0249212P.
PR   08-SEP-2000;  2000US-0231413P.         PR   17-NOV-2000;  2000US-0249213P.
PR   08-SEP-2000;  2000US-0231414P.         PR   17-NOV-2000;  2000US-0249214P.
PR   08-SEP-2000;  2000US-0232080P.         PR   17-NOV-2000;  2000US-0249215P.
PR   08-SEP-2000;  2000US-0232081P.         PR   17-NOV-2000;  2000US-0249216P.
PR   12-SEP-2000;  2000US-0211968P.         PR   17-NOV-2000;  2000US-0249217P.
PR   14-SEP-2000;  2000US-0232397P.         PR   17-NOV-2000;  2000US-0249218P.
PR   14-SEP-2000;  2000US-0232398P.         PR   17-NOV-2000;  2000US-0249244P.
PR   14-SEP-2000;  2000US-0232399P.         PR   17-NOV-2000;  2000US-0249245P.
PR   14-SEP-2000;  2000US-0232400P.         PR   17-NOV-2000;  2000US-0249264P.
PR   14-SEP-2000;  2000US-0232401P.         PR   17-NOV-2000;  2000US-0249265P.
PR   14-SEP-2000;  2000US-0233063P.         PR   17-NOV-2000;  2000US-0249297P.
PR   14-SEP-2000;  2000US-0233064P.         PR   17-NOV-2000;  2000US-0249299P.
PR   14-SEP-2000;  2000US-0233065P.         PR   17-NOV-2000;  2000US-0249300P.
PR   21-SEP-2000;  2000US-0234223P.         PR   01-DEC-2000;  2000US-0250160P.
PR   21-SEP-2000;  2000US-0234274P.         PR   05-DEC-2000;  2000US-0250391P.
PR   25-SEP-2000;  2000US-0234997P.         PR   05-DEC-2000;  2000US-0251030P.
                                            PR   05-DEC-2000;  2000US-0251988P.
                                            PR   06-DEC-2000;  2000US-0256719P.
                                            PR   08-DEC-2000;  2000US-0251479P.
                                            PR   08-DEC-2000;  2000US-0251856P.
                                            PR   08-DEC-2000;  2000US-0251868P.
                                            PR   08-DEC-2000;  2000US-0251869P.
                                            PR   08-DEC-2000;  2000US-0251989P.
                                            PR   11-DEC-2000;  2000US-0251990P.
                                            PR   05-JAN-2001;  2001US-0259678P.
```

```
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Rosen CA, Barash SC, Ruben SM;
XX
DR  WPI; 2001-476159/51.
DR  P-PSDB; ADM20145.
XX
PT  Isolated nucleic acid molecule encoding a channel/transporter protein is
PT  used in preventing, treating or ameliorating a medical condition.
XX
PS  Claim 1; SEQ ID NO 473; 809pp; English.
XX
CC  The invention relates to an isolated nucleic acid molecule encoding a
CC  channel/transporter protein or sequences at least 95% identical to a
CC  these. The nucleic acids and proteins encoded by them are used to
CC  prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC  rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC  in diagnosing a pathological condition or susceptibility to a
CC  pathological condition. The antibodies to the proteins can also be used
CC  in alleviating symptoms associated with the disorders and in diagnostic
CC  immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC  (ELISA). Disorders which are diagnosed or treated include autoimmune
CC  diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC  neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC  arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,
CC  nervous system disorders e.g. Alzheimer's disease, infections caused by
CC  bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC  The polypeptides-can also be used to aid wound healing and epithelial
CC  cell proliferation, to prevent skin aging due to sunburn, to maintain
CC  organs before transplantation, for supporting cell culture of primary
CC  tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC  also be used as a food additive or preservative to increase or decrease
CC  storage capabilities. This sequence corresponds to a gene of the
CC  invention.
XX
SQ  Sequence 549 BP; 126 A; 146 C; 128 G; 147 T; 0 U; 2 Other;
```

```
Alignment Scores:
Pred. No.:              154
Score:                  9.00
Percent Similarity:     100.0%           Length:          549
Best Local Similarity:  100.0%           Matches:         9
Query Match:            4.2%             Conservative:    0
DB:                                      Mismatches:      0
                                         Indels:          0
                                         Gaps:            0

US-09-989-890-238 (1-212) x ADM19666  (1-549)

Qy    24 LeuProSerSerLeuAlaLeuLeuSer 32
          |||||||||||||||||||||||||||
Db   349 TTGCCATCCTCTCTCGGCCTTGTTGTCC 375
```

```
RESULT 31
ADI54275
ID  ADI54275 standard; cDNA; 549 BP.
XX
AC  ADI54275;
XX
DT  22-APR-2004 (first entry)
XX
DE  cDNA encoding novel human protein seq id 478.
XX
KW  neuroprotective; nootropic; antiparkinsonian; anticonvulsant;
KW  antidiabetic; antirheumatic; antiarthritic; dermatological;
KW  antiinflammatory; immunosuppressive; antithyroid; antianemic; vasotropic;
KW  anti-HIV; hepatotropic; virucide; antibacterial; fungicide;
KW  antiparasitic; muscular; gynaecological; gastrointestinal; respiratory;
KW  cardiovascular-; antiarteriosclerotic; antiarrhythmic; cardiant;
KW  nephrotropic; litholytic; cytostatic; gene therapy; neural disorder;
KW  Alzheimer's disease; Parkinson's disease; Huntington's chorea;
KW  amyotrophic lateral sclerosis; multiple sclerosis;
KW  immune system disorder; diabetes; rheumatoid arthritis;
KW  systemic lupus erythematosus; autoimmune thyroiditis; haemolytic anaemia;
```

```
KW  inflammatory disorder; ischaemia-reperfusion injury;
KW  inflammatory bowel disease; Crohn's disease; infectious disease;
KW  HIV infection; hepatitis infection; bacterial infection;
KW  fungal infection; parasitic infection; muscular disorder;
KW  reproductive disorder; gastrointestinal disorder; pulmonary disorder;
KW  cardiovascular disorder; atherosclerosis; arrhythmia; myocarditis;
KW  renal disorder; acute glomerulonephritis; pyelonephritis;
KW  renal lithiasis; proliferative disorder; cancerous diseases; human; gene;
XX
OS  Homo sapiens.
XX
PN  US2004018969-A1.
XX
PD  29-JAN-2004.
XX
PF  17-JAN-2001; 2001US-00764875.
XX
PR  31-JAN-2000; 2000US-0179065P.
PR  04-FEB-2000; 2000US-0180628P.
PR  24-FEB-2000; 2000US-0184664P.
PR  02-MAR-2000; 2000US-0185350P.
PR  16-MAR-2000; 2000US-0189874P.
PR  17-MAR-2000; 2000US-0190076P.
PR  18-APR-2000; 2000US-0198123P.
PR  19-MAY-2000; 2000US-0205515P.
PR  07-JUN-2000; 2000US-0209467P.
PR  28-JUN-2000; 2000US-0214886P.
PR  30-JUN-2000; 2000US-0215135P.
PR  07-JUL-2000; 2000US-0216647P.
PR  07-JUL-2000; 2000US-0216880P.
PR  11-JUL-2000; 2000US-0217487P.
PR  11-JUL-2000; 2000US-0217496P.
PR  14-JUL-2000; 2000US-0218290P.
PR  26-JUL-2000; 2000US-0220963P.
PR  14-AUG-2000; 2000US-0220964P.
PR  14-AUG-2000; 2000US-0224518P.
PR  14-AUG-2000; 2000US-0224519P.
PR  14-AUG-2000; 2000US-0225213P.
PR  14-AUG-2000; 2000US-0225214P.
PR  14-AUG-2000; 2000US-0225266P.
PR  14-AUG-2000; 2000US-0225267P.
PR  14-AUG-2000; 2000US-0225268P.
PR  14-AUG-2000; 2000US-0225270P.
PR  14-AUG-2000; 2000US-0225447P.
PR  14-AUG-2000; 2000US-0225757P.
PR  14-AUG-2000; 2000US-0225758P.
PR  14-AUG-2000; 2000US-0225759P.
PR  18-AUG-2000; 2000US-0226279P.
PR  22-AUG-2000; 2000US-0226681P.
PR  22-AUG-2000; 2000US-0226868P.
PR  22-AUG-2000; 2000US-0227182P.
PR  23-AUG-2000; 2000US-0227009P.
PR  30-AUG-2000; 2000US-0229287P.
PR  01-SEP-2000; 2000US-0229343P.
PR  01-SEP-2000; 2000US-0229344P.
PR  01-SEP-2000; 2000US-0229345P.
PR  05-SEP-2000; 2000US-0229509P.
PR  05-SEP-2000; 2000US-0229513P.
PR  06-SEP-2000; 2000US-0230437P.
PR  06-SEP-2000; 2000US-0230438P.
PR  08-SEP-2000; 2000US-0231242P.
PR  08-SEP-2000; 2000US-0231243P.
PR  08-SEP-2000; 2000US-0231244P.
PR  08-SEP-2000; 2000US-0231413P.
PR  08-SEP-2000; 2000US-0231414P.
PR  08-SEP-2000; 2000US-0232080P.
PR  08-SEP-2000; 2000US-0232081P.
PR  12-SEP-2000; 2000US-0231968P.
PR  14-SEP-2000; 2000US-0232397P.
PR  14-SEP-2000; 2000US-0232398P.
PR  14-SEP-2000; 2000US-0232399P.
```

PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236377P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   01-NOV-2000; 2000US-0244474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.

PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2004-122079/12.
DR   P-PSDB; ADI54873.
XX
PT   New polypeptides and nucleic acid molecules, useful for detecting,
PT   preventing, diagnosing, prognosticating, treating or ameliorating medical
PT   conditions e.g. neural disorders, reproductive disorders or infectious
PT   diseases.
XX
PS   Claim 1; SEQ ID NO 478; 411pp; English.
XX
CC   The invention describes an isolated polypeptide comprising an amino acid
CC   sequence at least 90% identical to: a polypeptide fragment, domain,
CC   epitope, or full-length protein of any one of 607 amino acid sequences
CC   (I) described in the specification; a polypeptide fragment of (I), or the
CC   encoded sequence contained in (II), having biological activity; or a
CC   variant, allelic variant, or a species homologue of (I). The polypeptides
CC   and nucleic acid molecules are useful for detecting, preventing,
CC   diagnosing, prognosticating, treating or ameliorating medical conditions
CC   such as neural disorders, e.g. Alzheimer's disease, Parkinson's disease,
CC   Huntington's chorea, amyotrophic lateral sclerosis or multiple sclerosis,

Alignment Scores:
Pred. No.:   154   Length: 549
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x ADI54275 (1-549)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
         |||||||||||||||||||||||||||
Db   349 TTGCCATCCTCTCTGGCCTTGTTGTCC 375

RESULT 32
ABL89926
ID   ABL89926 standard; cDNA; 562 BP.
XX
AC   ABL89926;
XX
DT   24-MAY-2002 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 488.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-US016450.

XX
PR   19-MAY-2000; 2000US-0205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE, Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   P-PSDB; ABB89517.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders.
XX
PS   Claim 4; SEQ ID NO 488; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 562 BP; 126 A; 146 C; 134 G; 149 T; 0 U; 7 Other;

Alignment Scores:
Pred. No.:   157   Length: 562
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x ABL89926 (1-562)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
             |||||||||||||||||||||||||||
Db   355 TTGCCATCCTCTCTCGGCCTTGTTGTCC 381

RESULT 33
AAH98991
ID   AAH98991 standard; cDNA; 685 BP.
XX
AC   AAH98991;
XX
DT   12-OCT-2001 (first entry)
XX
DE   Rat EST-derived coding sequence SEQ ID NO: 848.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW   gene therapy; nutrition; ss.
XX
OS   Rattus norvegicus.
XX
PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US002687.

XX
PR   25-JAN-2000; 2000US-00491404.
PR   17-JUL-2000; 2000US-00617746.
PR   03-AUG-2000; 2000US-00631451.
PR   15-SEP-2000; 2000US-00663870.
XX
PA   (HYSB-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI   Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   P-PSDB; AAM24332.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use.
XX
PS   Claim 1; Page 701; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a cDNA
CC   of the invention
XX
SQ   Sequence 685 BP; 169 A; 178 C; 148 G; 190 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   188   Length: 685
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 4   Gaps: 0

US-09-989-890-238 (1-212) x AAH98991 (1-685)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
             |||||||||||||||||||||||||||
Db   648 TTGCCATCCTCTCTCGGCCTTGTTGTCC 674

RESULT 34
ADJ43042/c
ID   ADJ43042 standard; cDNA; 717 BP.
XX
AC   ADJ43042;
XX
DT   06-MAY-2004 (first entry)
XX
DE   Plant cDNA #4042.
XX
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
XX
OS   Eukaryota.
XX
PN   US2004016025-A1.
XX
PD   22-JAN-2004.
XX
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   26-SEP-2001; 2001US-0325448P.
PR   04-APR-2002; 2002US-0370620P.
XX

```
PA   (BUDW/)  BUDWORTH P.
PA   (MOUG/)  MOUGHAMER T.
PA   (BRIG/)  BRIGGS S P.
PA   (COOP/)  COOPER B.
PA   (GLAZ/)  GLAZEBROOK J.
PA   (GOFF/)  GOFF S A.
PA   (KATA/)  KATAGIRI F.
PA   (KREP/)  KREPS J.
PA   (PROV/)  PROVART N.
PA   (RICK/)  RICKE D.
PA   (ZHUT/)  ZHU T.
XX
PI   Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
PI   Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR   WPI; 2004-190374/18.
XX
PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Example 13; SEQ ID NO 4042; 230pp; English.
XX
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   have a high nutritional value with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 717 BP; 181 A; 161 C; 198 G; 177 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:             196                 Length:          717
Score:                 9.00                Matches:         9
Percent Similarity:    100.0%              Conservative:    0
Best Local Similarity: 100.0%              Mismatches:      0
Query Match:           4.2%                Indels:          0
DB:                    12                  Gaps:            0

US-09-989-890-238 (1-212) x ADJ43042 (1-717)

Qy    24 LeuProSerSerLeuAlaAlaLeuSer 32
         |||||||||||||||||||||||||||
Db   366 CTGCCTTCTTCATTGGCGCTGTTGTCG 340

RESULT 35
ADJ43040/c
ID   ADJ43040 standard; cDNA; 737 BP.
XX
AC   ADJ43040;
XX
DT   06-MAY-2004 (first entry)
XX
DE   Plant cDNA #4040.
XX
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
```

```
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
OS   Eukaryota.
XX
PN   US2004016025-A1.
XX
PD   22-JAN-2004.
XX
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   26-SEP-2001; 2001US-0325448P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (BUDW/)  BUDWORTH P.
PA   (MOUG/)  MOUGHAMER T.
PA   (BRIG/)  BRIGGS S P.
PA   (COOP/)  COOPER B.
PA   (GLAZ/)  GLAZEBROOK J.
PA   (GOFF/)  GOFF S A.
PA   (KATA/)  KATAGIRI F.
PA   (KREP/)  KREPS J.
PA   (PROV/)  PROVART N.
PA   (RICK/)  RICKE D.
PA   (ZHUT/)  ZHU T.
XX
PI   Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
PI   Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR   WPI; 2004-190374/18.
XX
PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Example 13; SEQ ID NO 4040; 230pp; English.
XX
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   have a high nutritional value with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 737 BP; 192 A; 162 C; 213 G; 170 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:             202                 Length:          737
Score:                 9.00                Matches:         9
Percent Similarity:    100.0%              Conservative:    0
Best Local Similarity: 100.0%              Mismatches:      0
Query Match:           4.2%                Indels:          0
DB:                    12                  Gaps:            0

US-09-989-890-238 (1-212) x ADJ43040 (1-737)

Qy    24 LeuProSerSerLeuAlaAlaLeuSer 32
         |||||||||||||||||||||||||||
Db   427 CTGCCTTCTTCGTTGGCGCTGTTGTCG 401
```

```
RESULT 36
AAH99445
ID   AAH99445 standard; cDNA; 792 BP.
XX
AC   AAH99445;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human protein encoding cDNA sequence SEQ ID NO:280.
XX
KW   Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW   antibacterial; endocrine; cardiant; central nervous system; virucide;
KW   anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW   antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW   dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW   neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW   immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW   antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW   cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW   genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW   thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW   allergic rhinitis; diabetes; multiple sclerosis; depression;
KW   Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW   neurological disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO20153455-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US035017.
XX
PR   23-DEC-1999;  99US-00471275.
PR   21-JAN-2000;  2000US-00488725.
PR   25-APR-2000;  2000US-00552317.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-457603/49.
DR   P-PSDB; AAM25504.
XX
PT   Isolated human polynucleotides encoding polypeptides, useful for the
PT   treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX
PS   Claim 1; Page 421; 1217pp; English.
XX
CC   AAH9166 to AAH99904 encode the human proteins given in AAM25225 to
CC   AAM25963. The proteins can have activities based on the tissues and cells
CC   they are expressed in, such as: activities/antibacterial; antirheumatic;
CC   antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC   central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC   cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
CC   antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC   antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC   antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC   encoding them can be used in gene therapy, antisense therapy and vaccine
CC   production. The proteins and polynucleotides are useful for screening for
CC   agonists or antagonists of a protein and for the treatment and diagnosis
CC   of disorders associated with the activity of a protein e.g. inflammation,
CC   rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC   neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC   infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC   anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC   osteoporosis, severe combined immunodeficiency, eczema, allergic
CC   rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC   Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC   neurological disorders
XX
```

```
SQ   Sequence 792 BP; 193 A; 201 C; 142 G; 256 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                    215          Length:              792
Score:                       9.00          Matches:               9
Percent Similarity:        100.0%          Conservative:          0
Best Local Similarity:     100.0%          Mismatches:            0
Query Match:                 4.2%          Indels:                0
DB:                                        Gaps:                  0

US-09-989-890-238 (1-212) x AAH99445 (1-792)

Qy         24 LeuProSerSerLeuAlaLeuLeuSer 32
              |||||||||||||||||||||||||||
Db        755 TTGCCATCCTCTCTGGCCTTGTTGTCC 781
```

```
RESULT 37
ADJ43041/c
ID   AD43041 standard; cDNA; 824 BP.
XX
AC   ADJ43041;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Plant cDNA #4041.
XX
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
XX
OS   Eukaryota.
XX
PN   US2004016025-A1.
XX
PD   22-JAN-2004.
XX
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2002; 2001US-0325277P.
PR   26-SEP-2002; 2001US-0325448P.
PR   04-APR-2002; 2002US-0376620P.
XX
PA   (BUDW/) BUDWORTH P.
PA   (MOUG/) MOUGHAMER T.
PA   (BRIG/) BRIGGS S P.
PA   (COOP/) COOPER B.
PA   (GLAZ/) GLAZEBROOK J.
PA   (GOFF/) GOFF S A.
PA   (KATA/) KATAGIRI F.
PA   (KREP/) KREPS J.
PA   (PROV/) PROVART N.
PA   (RICK/) RICKE D.
PA   (ZHUT/) ZHU T.
XX
PI   Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
PI   Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR   WPI; 2004-190374/18.
XX
PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Example 13; SEQ ID NO 4041; 230pp; English.
XX
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
```

CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC   have a high nutritional value with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 824 BP; 209 A; 183 C; 236 G; 196 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   223   Length: 824
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x ADJ43041 (1-824)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
         |||||||||||||||||||||||||||
Db   455 CTGCCTTCTTCATTGGCGCTGTTGTCG 429

RESULT 38
AD214553
ID   AD214553 standard; DNA; 1119 BP.
XX
AC   AD214553;
XX
DT   16-JUN-2005 (first entry)
XX
DE   DNA (LOC221103) encoding a human tumor associated antigen Seq 69.
XX
KW   chromosome 11; tumor-associated antigen; antisense therapy;
KW   RNA interference; diagnosis; cytostatic; cancer; metastasis; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2005030250-A2.
XX
PD   07-APR-2005.
XX
PF   23-SEP-2004; 2004WO-EP010697.
XX
PR   26-SEP-2003; 2003DE-01044799.
XX
PA   (GANY-) GANYMED PHARM AG.
XX
PI   Tuereci O, Sahin U, Helftenbein G, Schlueter V;
XX
DR   WPI; 2005-285105/29.
DR   P-PSDB; AD214554.
XX
PT   Compositions for treating and diagnosing cancer, contain agents that
PT   inhibit activity or expression of specific tumor-associated antigens, or
PT   bind to these antigens or nucleic acid encoding them.
XX
PS   Claim 1; SEQ ID NO 69; 388pp; German.
XX
CC   This invention relates to a novel pharmaceutical composition which
CC   comprises an agent that inhibits the activity or expression of a specific
CC   tumor-associated antigen (TAg). Specifically, it relates to tumor-
CC   associated antigens that are encoded by one of the following 75 nucleic
CC   acids sequences, fragments or derivatives thereof as given in the
CC   specification. The present invention describes antisense nucleic acids

CC   that hybridize to these TAg polynucleotides that may be used for
CC   antisense therapy and RNA interference, as well as methods for diagnosing
CC   a disease associated with (abnormal) expression of TAg. Accordingly, it
CC   further relates to methods for determining regression, progression and
CC   onset of a disease by administering an antibody, optionally linked to a
CC   therapeutic or diagnostic agent, that binds to TAg. As such, cytostatic
CC   compositions derived thereof are used for treating a wide range of
CC   cancers and their metastases, where the agents that bind specifically to
CC   TAg, and the nucleic acids that encode them, are useful for diagnosis and
CC   monitoring. This polynucleotide is a human DNA sequence encoding a tumor
CC   associated antigenic protein of the invention.
XX
SQ   Sequence 1119 BP; 318 A; 273 C; 209 G; 319 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   296   Length: 1119
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 14
DB: Gaps: 0

US-09-989-890-238 (1-212) x ADZ14553 (1-1119)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
         |||||||||||||||||||||||||||
Db   801 TTGCCATCCTCTCTCTGGCCTTGTTGTCC 827

RESULT 39
ABQ81557/c
ID   ABQ81557 standard; cDNA; 1209 BP.
XX
AC   ABQ81557;
XX
DT   30-DEC-2002 (first entry)
XX
DE   Gene up-regulated in metastatic colorectal cancer.
XX
KW   Colorectal cancer; metastasis; differential expression; cytostatic;
KW   diagnosis; gene therapy; vaccine; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1209
FT   /*tag= a
FT   /product= "Metastatic colon cancer polypeptide"
XX
PN   WO200268677-A2.
XX
PD   06-SEP-2002.
XX
PF   27-FEB-2002; 2002WO-US006001.
XX
PR   27-FEB-2001; 2001US-0272206P.
PR   02-APR-2001; 2001US-0281149P.
PR   17-APR-2001; 2001US-0284555P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
PA   (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI   Mack DH, Markowitz SD;
XX
DR   WPI; 2002-698677/75.
DR   P-PSDB; ABP54691.
XX
PT   New genes that are up- or down-regulated in colorectal cancer, useful for
PT   diagnosing colorectal cancer in a subject, or for identifying modulators
PT   of colorectal cancer-associated proteins and genes for treating
PT   colorectal cancer.
XX
PS   Claim 5; Page 252; 260pp; English.
XX

CC   The present sequence is the nucleotide sequence of a human gene that
CC   exhibits increased expression in metastatic colorectal cancer (MCC)
CC   samples. The gene is up-regulated in colon cancer-derived metastases
CC   compared to normal colon tissue. It is an example of claimed nucleic acid
CC   molecules that are up- or down-regulated in metastatic colorectal cancer
CC   cells. Such MCC-associated nucleic acids are useful in diagnostic and
CC   prognostic applications, in screening applications e.g. biochips, for
CC   identification of variant MCC-associated sequences, in informatics, for
CC   expression of MCC-associated proteins, in drug screening assays for
CC   identification of modulators of MCC-associated proteins, which are useful
CC   for treating this cancer, in gene therapy, as DNA vaccines, and as
CC   antisense or ribozyme modulators of MCC
XX
SQ   Sequence 1209 BP; 150 A; 485 C; 432 G; 141 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:   318   Length: 1209
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x ABQ81557 (1-1209)

Qy   59 AlaAlaAlaGlyIleAlaSerSerAla 67
            |||||||||||||||||||||||
Db   242 GCTGCTGCCGGGATCGCCTCTCCGCG 216

RESULT 40
ADL70222/c
ID   ADL70222 standard; cDNA; 1209 BP.
XX
AC   ADL70222;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Colon cancer marker ColoUp5 coding sequence.
XX
KW   ColoUp5; colon cancer; marker; diagnosis; cytostatic; vaccine;
KW   gene therapy; FoxQl; gene; ss.
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1209
FT   /*tag= a
FT   /product= "ColoUp5"
XX
PN   WO2004018648-A2.
XX
PD   04-MAR-2004.
XX
PF   26-AUG-2003; 2003WO-US027086.
XX
PR   26-AUG-2002; 2002US-00229345.
PR   27-AUG-2002; 2002US-0406296P.
PR   18-OCT-2002; 2002US-00274177.
XX
PA   (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI   Markowitz SD;
XX
DR   WPI; 2004-226834/21.
DR   P-PSDB; ADL70231.
XX
PT   Inhibiting the growth or proliferation of a colon neoplasia in a subject,
PT   for treating colon cancer, comprises administering to the subject an
PT   agent that decreases the amount of a polypeptide present in or produced
PT   by the colon neoplasia.
XX
PS   Claim 3; SEQ ID NO 9; 125pp; English.
XX

CC   The present sequence is the nucleic acid sequence of ColoUp5 a molecular
CC   marker of colon neoplasia. ColoUp5 was identified by expression
CC   microarray profiling of genes showing differential expression in normal
CC   colon and metastatic colon cancer. ColoUp5 is referred to in the
CC   literature as FoxQl (Forkhead box subclass q, member 1, formerly known as
CC   HFH-1), which plays a role in tissue-specific gene regulation and
CC   development, cell cycle regulation, cell signalling and tumourigenesis.
CC   The molecular marker is useful in the detection or diagnosis of colon
CC   neoplasia and for categorising the neoplastic state of a patient. Methods
CC   for inhibiting the growth or proliferation of a colon neoplasia involve
CC   the use of short interfering RNA (siRNA) or antisense oligonucleotides,
CC   or an agent that binds to and antagonises the polypeptide. Molecular
CC   markers can also be used to target therapeutic agents to cells of a colon
CC   neoplasia, and to screen and identify candidate therapeutic agents.
XX
SQ   Sequence 1209 BP; 150 A; 485 C; 432 G; 141 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:   318   Length: 1209
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x ADL70222 (1-1209)

Qy   59 AlaAlaAlaGlyIleAlaSerSerAla 67
            |||||||||||||||||||||||
Db   242 GCTGCTGCCGGGATCGCCTCTCCGCG 216

RESULT 41
ADN59614/c
ID   ADN59614 standard; DNA; 1209 BP.
XX
AC   ADN59614;
XX
DT   01-JUL-2004 (first entry)
XX
DE   Colon neoplasia ColoUp5 nucleic acid, SEQ ID No 9.
XX
KW   molecular marker; neoplastic state; colon neoplasia; ColoUp1; ColoUp2;
KW   colon cancer; ds; gene.
XX
OS   Unidentified.
XX
PN   WO2004018647-A2.
XX
PD   04-MAR-2004.
XX
PF   26-AUG-2003; 2003WO-US027085.
XX
PR   26-AUG-2002; 2002US-00229345.
PR   18-OCT-2002; 2002US-00274177.
XX
PA   (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI   Markowitz SD;
XX
DR   WPI; 2004-419536/39.
XX
PT   Detecting likeliness of subject to have colon neoplasia, involves
PT   obtaining biological sample from subject, detecting one or more of
PT   ColoUp1 and ColoUp2 polypeptides, presence of polypeptide indicating
PT   colon neoplasia.
XX
PS   Example 1; SEQ ID NO 9; 118pp; English.
XX
CC   The invention relates to novel molecular markers for categorizing the
CC   neoplastic state of a patient. The invention further comprises a method
CC   for detecting whether a subject is likely to have a colon neoplasia. The
CC   method comprises: obtaining a biological sample from the subject,
CC   detecting one or more polypeptides selected from among: one or more

```
CC   ColoUp1 polypeptides and one or more ColoUp2 polypeptides, wherein the
CC   presence of said one or more polypeptides is indicative of colon
CC   neoplasia. The invention further comprises a kit for carrying out the
CC   said method. The methods and compositions are useful for detecting colon
CC   cancer. This polynucleotide sequence represents the ColoUp5 gene of the
CC   invention.
XX
SQ   Sequence 1209 BP; 150 A; 485 C; 432 G; 141 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:                 318          Length:                1209
Score:                    9.00          Matches:                  9
Percent Similarity:      100.0%         Conservative:             0
Best Local Similarity:   100.0%         Mismatches:               0
Query Match:              4.2%          Indels:                   0
DB:                         12          Gaps:                     0

US-09-989-890-238 (1-212) x ADN59614 (1-1209)

Qy       59  AlaAlaAlaGlyIleAlaSerSerAla 67
             |||||||||||||||||||||||||||
Db      242  GCTGCTGCCGGGATCGCCTCCTCCGCG 216

RESULT 42
ACL71856/c
ID   ACL71856 standard; DNA; 1377 BP.
XX
AC   ACL71856;
XX
DT   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 2; SEQ ID NO 8319; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC   set of about 7842 genes or partial genes from the genome of the bacterium
CC   Mxyococcus xanthus. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from USPTO
XX
SQ   Sequence 1377 BP; 165 A; 420 C; 540 G; 252 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:                  359          Length:               1377
Score:                     9.00          Matches:                 9
Percent Similarity:       100.0%         Conservative:            0
Best Local Similarity:    100.0%         Mismatches:              0
Query Match:               4.2%          Indels:                  0
DB:                          14          Gaps:                    0

US-09-989-890-238 (1-212) x ACL71856 (1-1377),

Qy       56  SerAlaSerAlaAlaAlaGlyIleAla 64
             |||||||||||||||||||||||||||
Db     1289  AGTGCCAGCGGCGGCTGCGGAATCGCT 1263

RESULT 43
AAC52650/c
ID   AAC52650 standard; DNA; 1436 BP.
XX
AC   AAC52650;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 72008.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic pathway;
KW   promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   28-APR-1999;   99US-0130891P.
PR   30-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   14-MAY-1999;   99US-0134370P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0137782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137528P.
```

Fri Mar 17 09:43:20 2006

us-09-989-890-238.oligo_p2n.rng

```
PR  04-JUN-1999;  99US-0137502P.        PR  06-AUG-1999;  99US-0147303P.
PR  07-JUN-1999;  99US-0137724P.        PR  06-AUG-1999;  99US-0147416P.
PR  08-JUN-1999;  99US-0138094P.        PR  09-AUG-1999;  99US-0147493P.
PR  10-JUN-1999;  99US-0138540P.        PR  10-AUG-1999;  99US-0147935P.
PR  14-JUN-1999;  99US-0138847P.        PR  11-AUG-1999;  99US-0148171P.
PR  14-JUN-1999;  99US-0139119P.        PR  12-AUG-1999;  99US-0148319P.
PR  16-JUN-1999;  99US-0139452P.        PR  13-AUG-1999;  99US-0148341P.
PR  16-JUN-1999;  99US-0139453P.        PR  13-AUG-1999;  99US-0148565P.
PR  17-JUN-1999;  99US-0139492P.        PR  16-AUG-1999;  99US-0148684P.
PR  18-JUN-1999;  99US-0139454P.        PR  17-AUG-1999;  99US-0149368P.
PR  18-JUN-1999;  99US-0139455P.        PR  18-AUG-1999;  99US-0149175P.
PR  18-JUN-1999;  99US-0139456P.        PR  18-AUG-1999;  99US-0149426P.
PR  18-JUN-1999;  99US-0139457P.        PR  20-AUG-1999;  99US-0149722P.
PR  18-JUN-1999;  99US-0139458P.        PR  20-AUG-1999;  99US-0149723P.
PR  18-JUN-1999;  99US-0139459P.        PR  23-AUG-1999;  99US-0149929P.
PR  18-JUN-1999;  99US-0139460P.        PR  23-AUG-1999;  99US-0149902P.
PR  18-JUN-1999;  99US-0139461P.        PR  25-AUG-1999;  99US-0149930P.
PR  18-JUN-1999;  99US-0139462P.        PR  26-AUG-1999;  99US-0150566P.
PR  18-JUN-1999;  99US-0139463P.        PR  27-AUG-1999;  99US-0150884P.
PR  18-JUN-1999;  99US-0139750P.        PR  27-AUG-1999;  99US-0151065P.
PR  18-JUN-1999;  99US-0137763P.        PR  27-AUG-1999;  99US-0151066P.
PR  21-JUN-1999;  99US-0139817P.        PR  30-AUG-1999;  99US-0151080P.
PR  22-JUN-1999;  99US-0139899P.        PR  31-AUG-1999;  99US-0151303P.
PR  23-JUN-1999;  99US-0140353P.        PR  01-SEP-1999;  99US-0151438P.
PR  23-JUN-1999;  99US-0140354P.        PR  07-SEP-1999;  99US-0151930P.
PR  24-JUN-1999;  99US-0140695P.        PR  10-SEP-1999;  99US-0152363P.
PR  28-JUN-1999;  99US-0140823P.        PR  13-SEP-1999;  99US-0153070P.
PR  29-JUN-1999;  99US-0140991P.        PR  15-SEP-1999;  99US-0153758P.
PR  30-JUN-1999;  99US-0141287P.        PR  16-SEP-1999;  99US-0154018P.
PR  01-JUL-1999;  99US-0141842P.        PR  20-SEP-1999;  99US-0154039P.
PR  01-JUL-1999;  99US-0142154P.        PR  22-SEP-1999;  99US-0154779P.
PR  02-JUL-1999;  99US-0142055P.        PR  23-SEP-1999;  99US-0155139P.
PR  06-JUL-1999;  99US-0142390P.        PR  24-SEP-1999;  99US-0155486P.
PR  08-JUL-1999;  99US-0142803P.        PR  28-SEP-1999;  99US-0155659P.
PR  09-JUL-1999;  99US-0142920P.        PR  29-SEP-1999;  99US-0156458P.
PR  12-JUL-1999;  99US-0142977P.        PR  04-OCT-1999;  99US-0156596P.
PR  13-JUL-1999;  99US-0143542P.        PR  05-OCT-1999;  99US-0157117P.
PR  14-JUL-1999;  99US-0143624P.        PR  06-OCT-1999;  99US-0157753P.
PR  15-JUL-1999;  99US-0144005P.        PR  07-OCT-1999;  99US-0157865P.
PR  16-JUL-1999;  99US-0144085P.        PR  08-OCT-1999;  99US-0158029P.
PR  16-JUL-1999;  99US-0144086P.        PR  12-OCT-1999;  99US-0158232P.
PR  19-JUL-1999;  99US-0144325P.        PR  13-OCT-1999;  99US-0158369P.
PR  19-JUL-1999;  99US-0144331P.        PR  13-OCT-1999;  99US-0159293P.
PR  19-JUL-1999;  99US-0144332P.        PR  14-OCT-1999;  99US-0159294P.
PR  19-JUL-1999;  99US-0144333P.        PR  14-OCT-1999;  99US-0159295P.
PR  19-JUL-1999;  99US-0144334P.        PR  14-OCT-1999;  99US-0159329P.
PR  19-JUL-1999;  99US-0144335P.        PR  14-OCT-1999;  99US-0159330P.
PR  20-JUL-1999;  99US-0143352P.        PR  18-OCT-1999;  99US-0159331P.
PR  20-JUL-1999;  99US-0144632P.        PR  21-OCT-1999;  99US-0159637P.
PR  21-JUL-1999;  99US-0144884P.        PR  21-OCT-1999;  99US-0159638P.
PR  21-JUL-1999;  99US-0144814P.        PR  21-OCT-1999;  99US-0159584P.
PR  22-JUL-1999;  99US-0145086P.        PR  21-OCT-1999;  99US-0160741P.
PR  22-JUL-1999;  99US-0145088P.        PR  21-OCT-1999;  99US-0160767P.
PR  23-JUL-1999;  99US-0145087P.        PR  22-OCT-1999;  99US-0160768P.
PR  23-JUL-1999;  99US-0145089P.        PR  22-OCT-1999;  99US-0160770P.
PR  26-JUL-1999;  99US-0145276P.        PR  22-OCT-1999;  99US-0160814P.
PR  27-JUL-1999;  99US-0145913P.        PR  25-OCT-1999;  99US-0160815P.
PR  27-JUL-1999;  99US-0145918P.        PR  25-OCT-1999;  99US-0161405P.
PR  27-JUL-1999;  99US-0145919P.        PR  25-OCT-1999;  99US-0161406P.
PR  28-JUL-1999;  99US-0145951P.        PR  26-OCT-1999;  99US-0161359P.
PR  02-AUG-1999;  99US-0146386P.        PR  26-OCT-1999;  99US-0161360P.
PR  02-AUG-1999;  99US-0146388P.        PR  26-OCT-1999;  99US-0161361P.
PR  03-AUG-1999;  99US-0146389P.        PR  28-OCT-1999;  99US-0161920P.
PR  04-AUG-1999;  99US-0147038P.        PR  28-OCT-1999;  99US-0161992P.
PR  04-AUG-1999;  99US-0147204P.        PR  28-OCT-1999;  99US-0161993P.
PR  04-AUG-1999;  99US-0147302P.        PR  29-OCT-1999;  99US-0162142P.
PR  05-AUG-1999;  99US-0147192P.
PR  05-AUG-1999;  99US-0147260P.
```

Alignment Scores:
Pred. No.:   373   Length: 1436

```
Score:                      9.00      Matches:           9
Percent Similarity:       100.0%      Conservative:      0
Best Local Similarity:    100.0%      Mismatches:        0
Query Match:                4.2%      Indels:            0
DB:                            3      Gaps:              0

US-09-989-890-238 (1-212) x AACS2650 (1-1436)

Qy      130 LeuSerSerLysAlaSerLeuCysIle 138
            ||||||||||||||||||||||||
Db     1091 CTTTCCTCAAAGGCTAGCCTTTGTATC 1065

RESULT 44
ADM19426
ID    ADM19426 standard; cDNA; 1553 BP.
XX
AC    ADM19426;
XX
DT    20-MAY-2004  (first entry)
XX
DE    Novel human channel/transporter gene #223.
XX
KW    ds; gene; immunosuppressive; antiarthritic; antirheumatic;
KW    antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW    nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW    opthalmological; gene therapy; channel/transporter protein;
KW    rheumatoid arthritis; neoplasm; cardiac arrest; cerebrovascular disorder;
KW    cerebral ischemia; angiogenesis; nervous system disorder;
KW    Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW    epithelial cell proliferation; skin aging; sunburn; transplantation;
KW    chemotaxis; food additive.
XX
OS    Homo sapiens.
XX
PN    WO200154472-A2.
XX
PD    02-AUG-2001.
XX
PF    17-JAN-2001; 2001WO-US001307.
XX
PR    31-JAN-2000; 2000US-0179065P.
PR    04-FEB-2000; 2000US-0180628P.
PR    24-FEB-2000; 2000US-0184664P.
PR    02-MAR-2000; 2000US-0186350P.
PR    16-MAR-2000; 2000US-0189874P.
PR    17-MAR-2000; 2000US-0190076P.
PR    18-APR-2000; 2000US-0198123P.
PR    19-MAY-2000; 2000US-0205515P.
PR    07-JUN-2000; 2000US-0209467P.
PR    28-JUN-2000; 2000US-0214886P.
PR    30-JUN-2000; 2000US-0215135P.
PR    07-JUL-2000; 2000US-0216647P.
PR    07-JUL-2000; 2000US-0216880P.
PR    11-JUL-2000; 2000US-0217487P.
PR    11-JUL-2000; 2000US-0217496P.
PR    14-JUL-2000; 2000US-0218290P.
PR    26-JUL-2000; 2000US-0209963P.
PR    26-JUL-2000; 2000US-0209964P.
PR    14-AUG-2000; 2000US-0224518P.
PR    14-AUG-2000; 2000US-0224519P.
PR    14-AUG-2000; 2000US-0225213P.
PR    14-AUG-2000; 2000US-0225214P.
PR    14-AUG-2000; 2000US-0225266P.
PR    14-AUG-2000; 2000US-0225267P.
PR    14-AUG-2000; 2000US-0225268P.
PR    14-AUG-2000; 2000US-0225270P.
PR    14-AUG-2000; 2000US-0225447P.
PR    14-AUG-2000; 2000US-0257557P.
PR    14-AUG-2000; 2000US-0257558P.
PR    14-AUG-2000; 2000US-0257559P.
PR    18-AUG-2000; 2000US-0226279P.
PR    22-AUG-2000; 2000US-0226681P.
PR    22-AUG-2000; 2000US-0226868P.
PR    22-AUG-2000; 2000US-0227182P.
PR    23-AUG-2000; 2000US-0227009P.
PR    30-AUG-2000; 2000US-0228924P.
PR    01-SEP-2000; 2000US-0229287P.
PR    01-SEP-2000; 2000US-0229343P.
PR    01-SEP-2000; 2000US-0229344P.
PR    01-SEP-2000; 2000US-0229345P.
PR    05-SEP-2000; 2000US-0229509P.
PR    05-SEP-2000; 2000US-0229513P.
PR    06-SEP-2000; 2000US-0230437P.
PR    06-SEP-2000; 2000US-0230438P.
PR    08-SEP-2000; 2000US-0231242P.
PR    08-SEP-2000; 2000US-0231243P.
PR    08-SEP-2000; 2000US-0231244P.
PR    08-SEP-2000; 2000US-0231413P.
PR    08-SEP-2000; 2000US-0231414P.
PR    08-SEP-2000; 2000US-0232080P.
PR    12-SEP-2000; 2000US-0232081P.
PR    14-SEP-2000; 2000US-0231968P.
PR    14-SEP-2000; 2000US-0232397P.
PR    14-SEP-2000; 2000US-0232398P.
PR    14-SEP-2000; 2000US-0232399P.
PR    14-SEP-2000; 2000US-0232400P.
PR    14-SEP-2000; 2000US-0232401P.
PR    14-SEP-2000; 2000US-0233063P.
PR    14-SEP-2000; 2000US-0233064P.
PR    14-SEP-2000; 2000US-0233065P.
PR    21-SEP-2000; 2000US-0234223P.
PR    21-SEP-2000; 2000US-0234274P.
PR    25-SEP-2000; 2000US-0234997P.
PR    25-SEP-2000; 2000US-0234998P.
PR    26-SEP-2000; 2000US-0235484P.
PR    27-SEP-2000; 2000US-0235834P.
PR    27-SEP-2000; 2000US-0235836P.
PR    29-SEP-2000; 2000US-0236327P.
PR    29-SEP-2000; 2000US-0236367P.
PR    29-SEP-2000; 2000US-0236368P.
PR    29-SEP-2000; 2000US-0236369P.
PR    02-OCT-2000; 2000US-0236370P.
PR    02-OCT-2000; 2000US-0236802P.
PR    02-OCT-2000; 2000US-0237037P.
PR    02-OCT-2000; 2000US-0237038P.
PR    02-OCT-2000; 2000US-0237039P.
PR    02-OCT-2000; 2000US-0237040P.
PR    13-OCT-2000; 2000US-0239935P.
PR    13-OCT-2000; 2000US-0239937P.
PR    20-OCT-2000; 2000US-0240960P.
PR    20-OCT-2000; 2000US-0241221P.
PR    20-OCT-2000; 2000US-0241785P.
PR    20-OCT-2000; 2000US-0241786P.
PR    20-OCT-2000; 2000US-0241787P.
PR    20-OCT-2000; 2000US-0241808P.
PR    20-OCT-2000; 2000US-0241809P.
PR    20-OCT-2000; 2000US-0241826P.
PR    01-NOV-2000; 2000US-0244617P.
PR    08-NOV-2000; 2000US-0246474P.
PR    08-NOV-2000; 2000US-0246475P.
PR    08-NOV-2000; 2000US-0246476P.
PR    08-NOV-2000; 2000US-0246477P.
PR    08-NOV-2000; 2000US-0246478P.
PR    08-NOV-2000; 2000US-0246523P.
PR    08-NOV-2000; 2000US-0246524P.
PR    08-NOV-2000; 2000US-0246525P.
PR    08-NOV-2000; 2000US-0246526P.
PR    08-NOV-2000; 2000US-0246527P.
PR    08-NOV-2000; 2000US-0246528P.
PR    08-NOV-2000; 2000US-0246532P.
PR    08-NOV-2000; 2000US-0246609P.
PR    08-NOV-2000; 2000US-0246610P.
PR    08-NOV-2000; 2000US-0246611P.
PR    08-NOV-2000; 2000US-0246613P.
PR    17-NOV-2000; 2000US-0249207P.
PR    17-NOV-2000; 2000US-0249208P.
```

```
PR    17-NOV-2000;   2000US-0249209P.
PR    17-NOV-2000;   2000US-0249210P.
PR    17-NOV-2000;   2000US-0249211P.
PR    17-NOV-2000;   2000US-0249212P.
PR    17-NOV-2000;   2000US-0249213P.
PR    17-NOV-2000;   2000US-0249214P.
PR    17-NOV-2000;   2000US-0249215P.
PR    17-NOV-2000;   2000US-0249216P.
PR    17-NOV-2000;   2000US-0249217P.
PR    17-NOV-2000;   2000US-0249218P.
PR    17-NOV-2000;   2000US-0249244P.
PR    17-NOV-2000;   2000US-0249245P.
PR    17-NOV-2000;   2000US-0249264P.
PR    17-NOV-2000;   2000US-0249265P.
PR    17-NOV-2000;   2000US-0249297P.
PR    17-NOV-2000;   2000US-0249299P.
PR    17-NOV-2000;   2000US-0249300P.
PR    01-DEC-2000;   2000US-0250160P.
PR    01-DEC-2000;   2000US-0250391P.
PR    05-DEC-2000;   2000US-0251030P.
PR    05-DEC-2000;   2000US-0251888P.
PR    05-DEC-2000;   2000US-0256719P.
PR    06-DEC-2000;   2000US-0251479P.
PR    08-DEC-2000;   2000US-0258556P.
PR    08-DEC-2000;   2000US-0251868P.
PR    08-DEC-2000;   2000US-0258669P.
PR    08-DEC-2000;   2000US-0251989P.
PR    08-DEC-2000;   2000US-0251990P.
PR    11-DEC-2000;   2000US-0254097P.
PR    05-JAN-2001;   2001US-0259678P.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Rosen CA,  Barash SC,  Ruben SM;
XX
DR    WPI; 2001-476159/51.
DR    P-PSDB; ADM19905.
XX
PT    Isolated nucleic acid molecule encoding a channel/transporter protein is
PT    used in preventing, treating or ameliorating a medical condition.
XX
PS    Claim 1; SEQ ID NO 233; 809pp; English.
XX
CC    The invention relates to an isolated nucleic acid molecule encoding a
CC    channel/transporter protein or sequences at least 95% identical to a
CC    these. The nucleic acids and proteins encoded by them are used to
CC    prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC    rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC    in diagnosing a pathological condition or susceptibility to a
CC    pathological condition. The antibodies to the proteins can also be used
CC    in alleviating symptoms associated with the disorders and in diagnostic
CC    immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC    (ELISA). Disorders which are diagnosed or treated include autoimmune
CC    diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC    neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC    arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,
CC    nervous system disorders e.g. Alzheimer's disease, infections caused by
CC    bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC    The polypeptides can also be used to aid wound healing and epithelial
CC    cell proliferation, to prevent skin aging due to sunburn, to maintain
CC    organs before transplantation, for supporting cell culture of primary
CC    tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC    also be used as a food additive or preservative to increase or decrease
CC    storage capabilities. This sequence corresponds to a gene of the
CC    invention.
XX
SQ    Sequence 1553 BP; 480 A; 314 C; 313 G; 444 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:               401             Length:         1553
Score:                   9.00            Matches:        9
Percent Similarity:      100.0%          Conservative:   0
Best Local Similarity:   100.0%          Mismatches:     0
```

```
Query Match:        4.2%       Indels:       0
DB:                 5          Gaps:         0

US-09-989-890-238 (1-212) x ADM19426 (1-1553)

Qy      24  LeuProSerSerLeuAlaLeuLeuSer 32
            |||||||||||||||||||||||||||
Db     157  TTGCCATCCTCTCGGCCTTGTTGTCC 183

RESULT 45
ABK43589
ID   ABK43589 standard; cDNA; 1564 BP.
XX
AC   ABK43589;
XX
DT   05-JUN-2002  (first entry)
XX
DE   DNA encoding novel central nervous system protein #169.
XX
KW   Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW   cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW   acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW   adenocarcinoma; reproductive system disorder; testicular feminisation;
KW   endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW   respiratory disorder; renal disorder; kidney failure; blood disorder;
KW   myocardial infarction; wound healing; cell proliferation; skin aging;
KW   food additive; food preservative; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155318-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001332.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209063P.
PR   26-JUL-2000; 2000US-0209064P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0245519P.
PR   14-AUG-2000; 2000US-0255213P.
PR   14-AUG-2000; 2000US-0255214P.
PR   14-AUG-2000; 2000US-0255266P.
PR   14-AUG-2000; 2000US-0255267P.
PR   14-AUG-2000; 2000US-0255268P.
PR   14-AUG-2000; 2000US-0255270P.
PR   14-AUG-2000; 2000US-0255447P.
PR   14-AUG-2000; 2000US-0255757P.
PR   14-AUG-2000; 2000US-0255758P.
PR   14-AUG-2000; 2000US-0255759P.
PR   18-AUG-2000; 2000US-0262279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
```

```
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0304437P.
PR   08-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233061P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   02-OCT-2000;   2000US-0237040P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0246617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.

PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-581633/65.
DR   P-PSDB; AAU87259.
XX
PT   New isolated nucleic acid encoding a protein for diagnosing, preventing,
PT   treating or ameliorating medical conditions and used as food additives or
PT   preservatives.
XX
PS   Claim 1; SEQ ID NO 179; 837pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule (I) encoding a
CC   novel central nervous system protein. (I) and polypeptides (III) encoded
CC   by (I), are used to treat a medical conditions and in diagnosis of a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC   amylotrophic lateral sclerosis, infections caused by bacteria, viruses
CC   e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC   e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC   adenocarcinomas and irritable bowel syndrome, reproductive system
CC   disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC   and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC   leukaemia, disorders involving neovascularisation e.g. malignancies,
CC   respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC   acute kidney failure and blood related disorders e.g. myocardial
CC   infarction. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities, fat content, lipid, protein,
```

Alignment Scores:

Pred. No.:	403	Length:	1564
Score:	9.00	Matches:	9
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	4.2%	Indels:	0
DB:	4	Gaps:	0

```
US-09-989-890-238 (1-212) x ABK43589 (1-1564)

Qy      24 LeuProSerSerLeuAlaLeuLeuSer 32
           |||||||||||||||||||||||||||
Db     156 TTGCCATCCTCTCTGGCCTTGTTGTCC 182

RESULT 46
ADI53976
ID   ADI53976 standard; cDNA; 1564 BP.
XX
AC   ADI53976;
XX
DT   22-APR-2004  (first entry)
XX
DE   cDNA encoding novel human protein seq id 179.
XX
KW   neuroprotective; nootropic; antiparkinsonian; anticonvulsant;
KW   antidiabetic; antirheumatic; antiarthritic; dermatological;
KW   antiinflammatory; immunosuppressive; antithyroid; antianemic; vasotropic;
KW   anti-HIV; hepatotropic; virucide; antibacterial; fungicide;
KW   antiparasitic; muscular; gynaecological; gastrointestinal; respiratory;
KW   cardiovascular-; antiarteriosclerotic; antiarrhythmic; cardiant;
KW   nephrotropic; litholytic; cyostatic; gene therapy; neural disorder;
KW   Alzheimer's disease; Parkinson's disease; Huntington's chorea;
KW   amyotrophic lateral sclerosis; multiple sclerosis;
KW   immune system disorder; diabetes; rheumatoid arthritis;
KW   systemic lupus erythematosus; autoimmune thyroiditis; haemolytic anaemia;
KW   inflammatory disorder; ischaemia-reperfusion injury;
KW   inflammatory bowel disease; Crohn's disease; infectious disease;
KW   HIV infection; hepatitis infection; bacterial infection;
KW   fungal infection; parasitic infection; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; pulmonary disorder;
KW   cardiovascular disorder; atherosclerosis; arrhythmia; myocarditis;
KW   renal disorder; acute glomerulonephritis; pyelonephritis;
KW   renal lithiasis; proliferative disorder; cancerous diseases; human; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   US2004018969-A1.
XX
PD   29-JAN-2004.
XX
PF   17-JAN-2001; 2001US-00764875.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0209964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0254447P.
PR   14-AUG-2000; 2000US-0255757P.
PR   14-AUG-2000; 2000US-0255758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
```

```
PR    08-NOV-2000; 200US-0246609P.
PR    08-NOV-2000; 200US-0246610P.
PR    08-NOV-2000; 200US-0246611P.
PR    08-NOV-2000; 200US-0246613P.
PR    17-NOV-2000; 200US-0249207P.
PR    17-NOV-2000; 200US-0249208P.
PR    17-NOV-2000; 200US-0249209P.
PR    17-NOV-2000; 200US-0249210P.
PR    17-NOV-2000; 200US-0249211P.
PR    17-NOV-2000; 200US-0249212P.
PR    17-NOV-2000; 200US-0249213P.
PR    17-NOV-2000; 200US-0249214P.
PR    17-NOV-2000; 200US-0249215P.
PR    17-NOV-2000; 200US-0249216P.
PR    17-NOV-2000; 200US-0249217P.
PR    17-NOV-2000; 200US-0249218P.
PR    17-NOV-2000; 200US-0249244P.
PR    17-NOV-2000; 200US-0249245P.
PR    17-NOV-2000; 200US-0249264P.
PR    17-NOV-2000; 200US-0249265P.
PR    17-NOV-2000; 200US-0249297P.
PR    17-NOV-2000; 200US-0249299P.
PR    17-NOV-2000; 200US-0249300P.
PR    01-DEC-2000; 200US-0250160P.
PR    01-DEC-2000; 200US-0250391P.
PR    05-DEC-2000; 200US-0251030P.
PR    05-DEC-2000; 200US-0251988P.
PR    05-DEC-2000; 200US-0256719P.
PR    06-DEC-2000; 200US-0251479P.
PR    08-DEC-2000; 200US-0251856P.
PR    08-DEC-2000; 200US-0251868P.
PR    08-DEC-2000; 200US-0251869P.
PR    08-DEC-2000; 200US-0251989P.
PR    08-DEC-2000; 200US-0251990P.
PR    11-DEC-2000; 200US-0254097P.
PR    05-JAN-2001; 201US-0259678P.
XX
PA    (ROSE/) ROSEN C A.
PA    (RUBE/) RUBEN S M.
PA    (BARA/) BARASH S C.
XX
PI    Rosen CA, Ruben SM, Barash SC;
XX
DR    WPI; 2004-122079/12.
DR    P-PSDB; ADI54574.
XX
PT    New polypeptides and nucleic acid molecules, useful for detecting,
PT    preventing, diagnosing, prognosticating, treating or ameliorating medical
PT    conditions e.g. neural disorders, reproductive disorders or infectious
PT    diseases.
XX
PS    Claim 1; SEQ ID NO 179; 411pp; English.
XX
CC    The invention describes an isolated polypeptide comprising an amino acid
CC    sequence at least 90% identical to: a polypeptide fragment, domain,
CC    epitope, or full-length protein of any one of 607 amino acid sequences
CC    (I) described in the specification; a polypeptide fragment of (I), or the
CC    encoded sequence contained in (II), having biological activity; or a
CC    variant, allelic variant, or a species homologue of (I). The polypeptides
CC    and nucleic acid molecules are useful for detecting, preventing,
CC    diagnosing, prognosticating, treating or ameliorating medical conditions
CC    such as neural disorders, e.g. Alzheimer's disease, Parkinson's disease,
CC    Huntington's chorea, amyotrophic lateral sclerosis or multiple sclerosis,

Alignment Scores:
Pred. No.:              403            Length:         1564
Score:                  9.00           Matches:        9
Percent Similarity:     100.0%         Conservative:   0
Best Local Similarity:  100.0%         Mismatches:     0
Query Match:            4.2%           Indels:         0
DB:                     12             Gaps:           0

. US-09-989-890-238 (1-212) x ADI53976 (1-1564)
```

```
Qy          24 LeuProSerSerLeuAlaLeuLeuSer 32
               ||||||||||||||||||||||||||||
Db         156 TTGCCATCCTCTCTGGCCTTGTTGTCC 182

RESULT 47
ABS6295
ID    ABS6295 standard; cDNA; 1644 BP.
XX
AC    ABS6295;
XX
DT    17-JAN-2003  (first entry)
XX
DE    Human transporter protein cDNA.
XX
KW    Human; gene; ss; transporter protein; allelic variant;
KW    cell proliferation; cell differentiation; cell signalling; antibody;
KW    gene chip; transgenic; therapeutic; diagnostic.
XX
OS    Homo sapiens.
XX
FH    Key             Location/Qualifiers
FT    CDS             1..1644
FT                    /*tag= a
FT                    /product= "Transporter protein"
XX
PN    WO200279252-A1.
XX
PD    10-OCT-2002.
XX
PF    10-JAN-2002; 2002WO-US000456.
XX
PR    22-JAN-2001; 2001US-0262658P.
XX
PA    (PBKE ) PE CORP NY.
XX
PI    Merkulov G,  Beasley EM;
XX
DR    WPI; 2003-040658/03.
DR    P-PSDB; ABG71742.
XX
PT    New isolated human transporter proteins, useful for developing
PT    therapeutic or diagnostic compositions, particularly for developing human
PT    therapeutic agents that modulate transporter activity in cells or
PT    tissues.
XX
PS    Claim 4; Fig 1; 129pp; English.
XX
CC    The invention discloses an isolated human transporter polypeptide, and
CC    its allelic variants or orthologues, and the polynucleotides encoding
CC    them. Transporter proteins regulate many different functions of the cell,
CC    including cell proliferation, differentiation and signalling processes,
CC    by regulating the flow of molecules, such as ions and macromolecules,
CC    into and out of cells. The polynucleotide and polypeptide can be used to
CC    raise antibodies, create a gene chip, create a transgenic non-human
CC    animal, produce the novel polypeptide, detect the presence of the
CC    polypeptide or nucleic acid in a sample, identify a modulator of the
CC    polypeptide or its expression, identify a pharmaceutical composition, and
CC    carrier, that binds to the polypeptide and treat a disease or condition
CC    mediated by a human transporter protein which comprises administering to
CC    a patient the composition identified. The peptides and nucleic acid
CC    molecules are also useful in the development of human therapeutics and
CC    diagnostic compositions. The peptides are also useful for eliciting an
CC    immune response, to quantitatively determine levels of the protein (or
CC    its binding partner or ligand) in biological fluids or as markers for
CC    tissues in which the corresponding protein is preferentially expressed.
CC    The agents identified are useful for treating transporter-related
CC    conditions and a modulator of the peptide is also useful for treating a
CC    disorder characterised by an absence of, inappropriate or unwanted
CC    expression of the protein. The sequence presented is the human
CC    transporter protein cDNA
XX
SQ    Sequence 1644 BP; 389 A; 419 C; 374 G; 462 T; 0 U; 0 Other;
```

Alignment Scores:
Pred. No.:   422   Length: 1644
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 8

US-09-989-890-238 (1-212) x ABS55295 (1-1644)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
         |||||||||||||||||||||||||
Db 1511 TTGCCATCCTCTCTCGGCCTTGTTGTCC 1537

RESULT 48
ACL72571/c
ID ACL72571 standard; DNA; 1716 BP.
XX
AC ACL72571;
XX
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus gene sequence, seq id 9034.
XX
KW Transgenic plant; DNA replication; gene regulation; gene expression;
KW gene; ds.
XX
OS Myxococcus xanthus.
XX
PN US6833447-B1.
XX
PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR WPI; 2005-028716/03.
XX
PT New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
PS Example 2; SEQ ID NO 9034; 25pp; English.
XX
CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO.11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC set of about 7842 genes or partial genes from the genome of the bacterium
CC Myxococcus xanthus. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from USPTO
XX
SQ Sequence 1716 BP; 230 A; 580 C; 660 G; 246 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   439   Length: 1716
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 14

US-09-989-890-238 (1-212) x ACL72571 (1-1716)

Qy   50 SerGlyAlaGlyLeuProSerAlaSer 58
         |||||||||||||||||||||||||
Db 1342 TCAGGTGCTGGCCTTCCTTCAGCGAGT 1316

RESULT 49
AAL44695
ID AAL44695 standard;'cDNA; 1818 BP.
XX
AC AAL44695;
XX
DT 03-MAY-2002 (first entry)
XX
DE Human transporter and ion channel TRICH-29 cDNA.
XX
KW Human; transporter and ion channel; TRICH; transport disorder;
KW neurological disorder; muscle disorder; immunological disorder;
KW cell proliferative disorder; neuroprotective; nootropic;
KW cerebroprotective; immunosuppressive; cytostatic; respiratory; muscular;
KW gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO20204520-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021448.
XX
PR 07-JUL-2000; 2000US-0216547P.
PR 14-JUL-2000; 2000US-0218232P.
PR 21-JUL-2000; 2000US-0220112P.
PR 28-JUL-2000; 2000US-0221839P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Raumann BE, Thornton M, Ding L, Yue H, Tang YT, Harland L;
PI Burford N, Greene BD, Sanjanwala MS, Baughn MR, Yao MG, Yang J;
PI Patterson C, Gandhi AR, Hafalia AJA, Tribouley CM, Walia NK;
PI Au-Young J, Walsh RT, Ramkumar J, Lu Y, Lu DAM, Azimzai Y, Lal P;
PI Elliott VS, Nguyen DB, Xu Y, Seilhamer JJ, Borowsky ML, Khan FA;
PI Kearney L, Thangavelu K, Das D, Policky JL;
XX
DR WPI; 2002-205969/26.
DR P-PSDB; AAO14212.
XX
PT New human transporters and ion channel polypeptides for diagnosing,
PT treating or preventing transport, neurological, muscle, immunological and
PT cell proliferative disorders.
XX
PS Claim 5; Page 227; 230pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human transporter and ion channel proteins, designated TRICH-1-
CC TRICH-32. The sequences can be used in the treatment of transport,
CC neurological, muscle, immunological and cell proliferative disorders. The
CC present sequence is a coding sequence of the invention
XX
SQ Sequence 1818 BP; 435 A; 455 C; 417 G; 511 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   463   Length: 1818
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 6

US-09-989-890-238 (1-212) x AAL44695 (1-1818)

Qy   24 LeuProSerLeuAlaLeuLeuSer 32
         |||||||||||||||||||||||

Db   1685 TTGCCATCCTCTCTGGCCTGTTGTCC 1711

RESULT 50
ADR86567
ID   ADR86567 standard; cDNA; 1912 BP.
XX
AC   ADR86567;
XX
DT   16-DEC-2004 (first entry)
XX
DE   UST3-LIKE1 protein encoding sequence.
XX
KW   UST3-LIKE1; Hepatotropic; Antiinflammatory; Cytostatic; Antidiabetic;
KW   Cardiant; Hemostatic; Vasotropic; Osteopathic; Antiparkinsonian;
KW   Antianemic; Laxative; Antidiarrheic; Antiasthmatic; Antiallergic;
KW   Cerebroprotective; antiarrhythmic; gastrointestinal; cirrhosis; tumour;
KW   Alzheimer's disease; Parkinson's disease; anemia; leukemia; lymphomas;
KW   asthma; dementia; arrhythmias; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   61..1750
FT   /*tag= a
FT   /product= "UST3-LIKE1"
XX
PN   WO2004081041-A1.
XX
PD   23-SEP-2004.
XX
PF   04-MAR-2004; 2004WO-EP002172.
XX
PR   10-MAR-2003; 2003EP-00004939.
XX
PA   (FARB) BAYER HEALTHCARE AG.
XX
PI   Golz S,   Brueggemeier U,   Geerts A;
XX
DR   WPI; 2004-677521/66.
DR   P-PSDB; ADR86568.
XX
PT   New nucleic acid molecule encoding an organic anion transporting (OAT)-
PT   like protein UST3-LIKE1, useful for treating gastrointestinal, metabolic,
PT   respiratory, neurological and cardiovascular diseases.
XX
PS   Claim 1; SEQ ID NO 1; 130pp; English.
XX
CC   The present invention relates to a nucleic acid molecule encoding a UST3-
CC   LIKE1. The UST3-LIKE1 or its regulator is useful for the treatment of
CC   gastrointestinal and liver diseases, metabolic diseases, hematological
CC   disorders, respiratory diseases, neurological disorders, urological
CC   disorders and cardiovascular diseases in a mammal. The regulator is a
CC   small molecule, an RNA molecule, an antisense oligonucleotide, a
CC   polypeptide, an antibody, or a ribozyme. The disorders include liver
CC   cirrhosis, liver tumours, diarrhea, achalasia, dysphagia, gastroesophageal reflux,
CC   diabetes, Crohn's disease, diarrhea, constipation, osteoporosis,
CC   Alzheimer's disease, Parkinson's disease, anemia, leukemia, lymphomas,
CC   asthma, allergies, stroke, dementia and arrhythmias. The present sequence
CC   represents UST3-LIKE1 protein encoding sequence.
XX
SQ   Sequence 1912 BP; 495 A; 461 C; 413 G; 543 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   485   Length: 1912
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 13   Gaps: 0

US-09-989-890-238 (1-212) x ADR86567 (1-1912)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32

Db   1589 TTGCCATCCTCTCTGGCCTTGTTGTCC 1615

RESULT 51
ADN33172
ID   ADN33172 standard; DNA; 1960 BP.
XX
AC   ADN33172;
XX
DT   18-NOV-2004 (first entry)
XX
DE   Human transporter and ion channel (TRICH) gene SeqID57.
XX
KW   transporters and ion channel; TRICH; neuroprotective; relaxant;
KW   antithyroid; antidiabetic; cytostatic; dermatological; immunosuppressive;
KW   antiinflammatory; thyromimetic; antiallergic; cerebroprotective;
KW   gastrointestinal; hepatotropic; nephrotropic; anticonvulsant;
KW   antiparkinsonian; antibacterial; antiparasitic; fungicide; protozoacide;
KW   virucide; uropathic; antirheumatic; cardiant; cardiovascular; anti-HIV;
KW   nootropic; TRICH agonist; TRICH antagonist; gene therapy.
KW   neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
KW   muscular disorder; myotonic dystrophy; catatonia; endocrine disorder;
KW   diabetes; Grave's disease; cancer; leukaemia; cervical cancer;
KW   breast cancer; immunological disorder; scleroderma;
KW   systemic lupus erythematosus; allergy; gastrointestinal disorder;
KW   Crohn's disease; renal disorder; Goodpasture's syndrome; infection;
KW   viral; bacterial; fungal; parasitic; protozoal; helminthic;
KW   cardiovascular disorder; atherosclerosis; hepatic disease; cirrhosis;
KW   transgenic animal; gene; ds; human.
XX
OS   Homo sapiens.
XX
PN   WO2004035755-A2.
XX
PD   29-APR-2004.
XX
PF   16-OCT-2003; 2003WO-US033087.
XX
PR   16-OCT-2002; 2002US-0419313P.
PR   23-OCT-2002; 2002US-0421033P.
PR   25-OCT-2002; 2002US-0421349P.
PR   04-NOV-2002; 2002US-0423516P.
XX
PA   (INCY-) INCYTE CORP.
XX
PI   Hafalia AJA,   Khare R,   Lal PG,   Yue H,   Baughn MR,   Thornton MB;
PI   Lu DAM,   Ison CH,   Becha SD,   Ding L,   Warren BA,   Lee SY,   Swarnakar A;
PI   Elliott VS,   Richardson TW,   Marquis JP,   Ramkumar J,   Murage J;
PI   Raumann BE,   Yao MG,   Lu Y,   Gietzen KJ,   Yang YG,   Chang H,   Chawla NK;
PI   Tran UK,   Lee S,   Yang J,   Gandhi AR,   Tribouley CM,   Policky JL;
PI   Ring HZ,   Lee EA;
XX
DR   WPI; 2004-348448/32.
DR   P-PSDB; ADN33116.
XX
PT   New TRICH polypeptides, useful for diagnosing, preventing, and treating
PT   disorders associated with abnormal expression or activity of TRICH, e.g.
PT   neuromuscular, immunological, cardiovascular disorders, cancer and/or
PT   infections.
XX
PS   Claim 5; SEQ ID NO 57; 285pp; English.
XX
CC   This invention relates to the nucleotide sequences which encode them. The
CC   (TRICH) proteins and the nucleotide sequences which encode them. The
CC   invention may be useful for the production of compounds with a
CC   neuroprotective, relaxant, antithyroid, antidiabetic, cytostatic,
CC   dermatological, immunosuppressive, antiinflammatory, thyromimetic,
CC   antiallergic, cerebroprotective, gastrointestinal, hepatotropic,
CC   nephrotropic, anticonvulsant, antiparkinsonian, antibacterial,
CC   antiparasitic, fungicide, protozoacide, virucide, uropathic,
CC   antirheumatic, cardiant, cardiovascular, anti-HIV or nootropic activity
CC   acting as TRICH agonists or antagonists. In addition the disclosed
CC   sequences may be useful for gene therapy. The invention may be useful in

CC   diagnosing, preventing, and treating disorders associated with an
CC   abnormal expression or activity of TRICH, such as neurodegenerative
CC   disorders (for example Parkinson's disease, Alzheimer's disease),
CC   muscular disorders (for example myotonic dystrophy, catatonia), endocrine
CC   disorders (for example diabetes, Grave's disease), cancers (for example
CC   leukaemia, cervical or breast cancers), immunological disorders (for
CC   example scleroderma, systemic lupus erythematosus, allergies),
CC   gastrointestinal disorders (for example Crohn's disease), renal disorders
CC   (for example Goodpasture's syndrome), infections (for example viral,
CC   bacterial, fungal, parasitic, protozoal, helminthic), cardiovascular
CC   disorders (for example atherosclerosis), or hepatic diseases (for example
CC   cirrhosis). TRICH or its fragments may also be used in screening for
CC   compounds that specifically bind to and modulate the activity of TRICH.
CC   The nucleotides can be used to model human disease. The present sequence
CC   animals to create humanised animals or transgenic
CC   transporter and ion channel (TRICH) gene of the invention.
XX
SQ   Sequence 1960 BP; 457 A; 512 C; 448 G; 543 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   496   Length: 1960
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 13   Gaps: 0

US-09-989-890-238 (1-212) x ADN33172 (1-1960)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
         ||||||||||||||||||||||||
Db   1800 TTGCCATCCTCCTCTGGCCTTGTTGTCC 1826

RESULT 52
ADS47648/c
ID   ADS47648 standard; cDNA; 2071 BP.
XX
AC   ADS47648;
XX
DT   02-DEC-2004 (first entry)
XX
DE   Bacterial polynucleotide #2391.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polynucleotide; gene; ss.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a

---

PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 26078; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polynucleotide used in
CC   the scope of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 2071 BP; 608 A; 350 C; 429 G; 684 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   522   Length: 2071
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 13   Gaps: 0

US-09-989-890-238 (1-212) x ADS47648 (1-2071)

Qy   175 ValLeuValAlaProArgSerThrIle 183
         ||||||||||||||||||||||||||
Db   1572 GTCTTAGTAGCCCTCGAAGCACGATT 1546

RESULT 53
ADN33202
ID   ADN33202 standard; DNA; 2238 BP.
XX
AC   ADN33202;
XX
DT   18-NOV-2004 (first entry)
XX
DE   Human transporter and ion channel (TRICH) gene SeqID87.
XX
KW   transporters and ion channel; TRICH; neuroprotective; relaxant;
KW   antithyroid; antidiabetic; cytostatic; dermatological; immunosuppressive;
KW   antiinflammatory; thyromimetic; antiallergic; cerebroprotective;
KW   gastrointestinal; hepatotropic; nephrotropic; anticonvulsant;
KW   antiparkinsonian; antibacterial; antiparasitic; fungicide; protozoacide;
KW   virucide; uropathic; antirheumatic; cardiant; cardiovascular; anti-HIV;
KW   nootropic; TRICH agonist; TRICH antagonist; gene therapy;
KW   neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
KW   muscular disorder; myotonic dystrophy; catatonia; endocrine disorder;
KW   diabetes; Grave's disease; cancer; leukaemia; cervical cancer;
KW   breast cancer; immunological disorder; scleroderma;
KW   systemic lupus erythematosus; allergy; gastrointestinal disorder;
KW   Crohn's disease; renal disorder; Goodpasture's syndrome; infection;
KW   viral; bacterial; fungal; parasitic; protozoal; helminthic;
KW   cardiovascular disorder; atherosclerosis; hepatic disease; cirrhosis;
KW   transgenic animal; gene; ds; human.
XX
OS   Homo sapiens.

XX
PN   WO2004035755-A2.
XX
PD   29-APR-2004.
XX
PF   16-OCT-2003; 2003WO-US033087.
XX
PR   16-OCT-2002; 2002US-0419313P.
PR   23-OCT-2002; 2002US-0421033P.
PR   25-OCT-2002; 2002US-0421349P.
PR   04-NOV-2002; 2002US-0423516P.
XX
PA   (INCY-) INCYTE CORP.
XX
PI   Hafalia AJA, Khare R, Lal PG, Yue H, Baughn MR, Thornton MB;
PI   Lu DAM, Ison CH, Becha SD, Ding L, Warren BA, Lee SY, Swarnakar A;
PI   Elliott VS, Richardson TW, Marquis JP, Ramkumar J, Murage J;
PI   Raumann BE, Yao MG, Lu Y, Gietzen KJ, Yang YG, Chang H, Chawla NK;
PI   Tran UK, Lee S, Yang J, Gandhi AR, Tribouley CM, Policky JL;
PI   Ring HZ, Lee EA;
XX
DR   WPI; 2004-348448/32.
DR   P-PSDB; ADN33146.
XX
PT   New TRICH polypeptides, useful for diagnosing, preventing, and treating
PT   disorders associated with abnormal expression or activity of TRICH, e.g.
PT   neuromuscular, immunological, cardiovascular disorders, cancer and/or
PT   infections.
XX
PS   Claim 5; SEQ ID NO 87; 285pp; English.
XX
CC   This invention relates to novel human transporters and ion channel
CC   (TRICH) proteins and the nucleotide sequences which encode them. The
CC   invention may be useful for the production of compounds with a
CC   neuroprotective, relaxant, antithyroid, antidiabetic, cytostatic,
CC   dermatological, immunosuppressive, antiinflammatory, thyromimetic,
CC   antiallergic, cerebroprotective, gastrointestinal, hepatotropic,
CC   nephrotropic, anticonvulsant, antiparkinsonian, antibacterial,
CC   antiparasitic, fungicide, protozoacide, virucide, uropathic,
CC   antirheumatic, cardiant, cardiovascular, anti-HIV or nootropic activity
CC   acting as TRICH agonists or antagonists. In addition the disclosed
CC   sequences may be useful for gene therapy. The invention may be useful in
CC   diagnosing, preventing, and treating disorders associated with an
CC   abnormal expression or activity of TRICH, such as neurodegenerative
CC   disorders (for example Parkinson's disease, Alzheimer's disease),
CC   muscular disorders (for example myotonic dystrophy, catatonia), endocrine
CC   disorders (for example diabetes, Grave's disease), cancers (for example
CC   leukaemia, cervical or breast cancers), immunological disorders (for
CC   example scleroderma, systemic lupus erythematosus, allergies),
CC   gastrointestinal disorders (for example Crohn's disease), renal disorders
CC   (for example Goodpasture's syndrome), infections (for example viral,
CC   bacterial, fungal, parasitic, protozoal, helminthic), cardiovascular
CC   disorders (for example atherosclerosis), or hepatic diseases (for example
CC   cirrhosis). TRICH or its fragments may also be used in screening for
CC   compounds that specifically bind to and modulate the activity of TRICH.
CC   The nucleotides can be used to create humanised animals or transgenic
CC   animals to model human disease. The present sequence is that of a human
CC   transporter and ion channel (TRICH) gene of the invention.
XX
SQ   Sequence 2238 BP; 547 A; 563 C; 506 G; 622 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   561   Length: 2238
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x ADN33202 (1-2238)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
              |||||||||||||||||||||||||||

Db   2078 TTGCCATCCTCTCTCGGCCTTGTGTCC 2104

RESULT 54
AAL44696
ID   AAL44696 standard; cDNA; 2245 BP.
XX
AC   AAL44696;
XX
DT   03-MAY-2002 (first entry)
XX
DE   Human transporter and ion channel TRICH-30 cDNA.
XX
KW   Human; transporter and ion channel; TRICH; transport disorder;
KW   neurological disorder; muscle disorder; immunological disorder;
KW   cell proliferative disorder; neuroprotective; nootropic;
KW   cerebroprotective; immunosuppressive; cytostatic; respiratory; muscular;
KW   gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200204520-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021448.
XX
PR   07-JUL-2000; 2000US-0216547P.
PR   14-JUL-2000; 2000US-0218232P.
PR   21-JUL-2000; 2000US-0220112P.
PR   28-JUL-2000; 2000US-0221839P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Raumann BE, Thornton M, Ding L, Yue H, Tang YT, Harland L;
PI   Burford N, Greene BD, Sanjanwala MS, Baughn MR, Yao MG, Yang J;
PI   Patterson C, Gandhi AR, Hafalia AJA, Tribouley CM, Walia NK;
PI   Au-Young J, Walsh RT, Ramkumar J, Lu Y, Lu DAM, Azimzai Y, Lal P;
PI   Elliott VS, Nguyen DB, Xu Y, Seilhamer JJ, Borowsky ML, Khan FA;
PI   Kearney L, Thangavelu K, Das D, Policky JL;
XX
DR   WPI; 2002-205969/26.
DR   P-PSDB; AA014213.
XX
PT   New human transporters and ion channel polypeptides for diagnosing,
PT   treating or preventing transport, neurological, muscle, immunological and
PT   cell proliferative disorders.
XX
PS   Claim 5; Page 228; 230pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human transporter and ion channel proteins, designated TRICH-1-
CC   TRICH-32. The sequences can be used in the treatment of transport,
CC   neurological, muscle, immunological and cell proliferative disorders. The
CC   present sequence is a coding sequence of the invention
XX
SQ   Sequence 2245 BP; 550 A; 559 C; 491 G; 645 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 2245
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x AAL44696 (1-2245)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
              |||||||||||||||||||||||||||
Db   2112 TTGCCATCCTCTCTCGGCCTTGTGTCC 2138

RESULT 55
AAS79799/c

Column 1

```
ID    AAS79799 standard; cDNA; 2256 BP.
XX
AC    AAS79799;
XX
DT    13-FEB-2002  (first entry)
XX
DE    DNA encoding novel human diagnostic protein #15603.
XX
KW    Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW    food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS    Homo sapiens.
XX
PN    WO200175067-A2.
XX
PD    11-OCT-2001.
XX
PF    30-MAR-2001; 2001WO-US008631.
XX
PR    31-MAR-2000; 2000US-00540217.
PR    23-AUG-2000; 2000US-00649167.
XX
PA    (HYSB-) HYSEQ INC.
XX
PI    Drmanac RT, Liu C, Tang YT;
XX
DR    WPI; 2001-639362/73.
DR    P-PSDB; ABG15612.
XX
PT    New isolated polynucleotide and encoded polypeptides, useful in
PT    diagnostics, forensics, gene mapping, identification of mutations
PT    responsible for genetic disorders or other traits and to assess
PT    biodiversity.
XX
PS    Claim 1; SEQ ID NO 15603; 103pp; English.
XX
CC    The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC    sequences. (I) is useful as hybridisation probes, polymerase chain
CC    reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC    and in recombinant production of (II). The polynucleotides are also used
CC    in diagnostics as expressed sequence tags for identifying expressed
CC    genes. (I) is useful in gene therapy techniques to restore normal
CC    activity of (II) or to treat disease states involving (II). (II) is
CC    useful for generating antibodies against it, detecting or quantitating a
CC    polypeptide in tissue, as molecular weight markers and as a food
CC    supplement. (II) and its binding partners are useful in medical imaging
CC    of sites expressing (II). (I) and (II) are useful for treating disorders
CC    involving aberrant protein expression or biological activity. The
CC    polypeptide and polynucleotide sequences have application in
CC    diagnostics, forensics, gene mapping, identification of mutations
CC    responsible for genetic disorders or other traits to assess biodiversity
CC    and to produce other types of data and products dependent on DNA and
CC    amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC    coding sequences of the invention. Note: The sequence data for this
CC    patent did not appear in the printed specification, but was obtained in
CC    electronic format directly from WIPO at
CC    ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 2256 BP; 487 A; 771 C; 609 G; 389 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                 565           Length:              2256
Score:                     9.00          Matches:                9
Percent Similarity:        100.0%        Conservative:           0
Best Local Similarity:     100.0%        Mismatches:             0
Query Match:               4.2%          Indels:                 0
DB:                        5             Gaps:                   0

US-09-989-890-238 (1-212) x AAS79799 (1-2256)

Qy    54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
         |||||||||||||||||||||||||||||||
Db    337 CTTCCTTCAGCGTCAGCAGCAGCTGGA 311
```

Column 2

```
RESULT 56
ACN43045/c
ID    ACN43045 standard; cDNA; 2275 BP.
XX
AC    ACN43045;
XX
DT    18-NOV-2004  (first entry)
XX
DE    Human diagnostic and therapeutic polynucleotide SEQ ID NO:1920.
XX
KW    ss; gene; gene therapy; human diagnostic and therapeutic polynucleotide;
KW    dithp.
XX
OS    Homo sapiens.
XX
PN    WO2004023973-A2.
XX
PD    25-MAR-2004.
XX
PF    12-SEP-2003; 2003WO-US028227.
XX
PR    12-SEP-2002; 2002US-0410259P.
PR    12-SEP-2002; 2002US-0410260P.
XX
PA    (INCY-) INCYTE CORP.
XX
PI    Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI    Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI    Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI    Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI    Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI    Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI    Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI    Patury S, Shi X, Suarez CJ;
XX
DR    WPI; 2004-329368/30.
DR    P-PSDB; ABM84393.
XX
PT    New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT    in diagnosing a condition, disease or disorder associated with human
PT    molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT    in gene mapping.
XX
PS    Claim 1; Page; 190pp; English.
XX
CC    The invention relates to novel diagnostic and therapeutic polynucleotides
CC    selected from one of the 2722 sequences defined in the specification. A
CC    polynucleotide of the invention may have a use in gene therapy. The human
CC    diagnostic and therapeutic polynucleotide (dithp) or polypeptides may be
CC    used to diagnose a particular condition, disease or disorder associated
CC    with human molecules, e.g. cell proliferative disorders,
CC    autoimmune/inflammatory disorder, developmental disorder, endocrine
CC    disorder, neurological disorders, gastrointestinal disorders, or
CC    infections caused by virus, bacteria, fungi or parasite. The dithp
CC    molecules may also be used in genetic mapping, in identifying individuals
CC    from minute biological samples, in detecting single nucleotide
CC    polymorphisms, as molecular weight markers, and for somatic or germline
CC    gene therapy. The present sequence represents a dithp polynucleotide of
CC    the invention. Note: The sequence data for this patent is not represented
CC    in the printed specification, but was obtained in electronic format
CC    directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ    Sequence 2275 BP; 472 A; 691 C; 624 G; 488 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                 570           Length:              2275
Score:                     9.00          Matches:                9
Percent Similarity:        100.0%        Conservative:           0
Best Local Similarity:     100.0%        Mismatches:             0
Query Match:               4.2%          Indels:                 0
DB:                        13            Gaps:                   0
```

US-09-989-890-238 (1-212) x ACN43045 (1-2275)

Qy   25 ProSerLeuAlaLeuLeuSerArg 33
           ||||||||||||||||||||
Db   1955 CCGAGCTCCCTGGCTCTACTATCCAGG 1929

RESULT 57
ABL17805
ID   ABL17805 standard; DNA; 3840 BP.
XX
AC   ABL17805;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 4888.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 4888; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB5737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3840 BP; 985 A; 1179 C; 1029 G; 647 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   923   Length: 3840
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 4   Gaps: 0

US-09-989-890-238 (1-212) x ABL17805 (1-3840)

Qy   54 LeuProSerAlaSerAlaAlaAlaGly 62
           |||||||||||||||||||||||||||
Db   567 CTACCATCCGCCAGCGCCGCCGCCGGT 593

RESULT 58
ABL17804/c
ID   ABL17804 standard; DNA; 5906 BP.
XX
AC   ABL17804;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 4885.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 4885; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB5737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5906 BP; 1234 A; 1501 C; 1649 G; 1522 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   1.37e+03   Length: 5906
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 4   Gaps: 0

US-09-989-890-238 (1-212) x ABL17804 (1-5906)

Qy   54 LeuProSerAlaSerAlaAlaAlaGly 62
           |||||||||||||||||||||||||||
Db   4274 CTACCATCCGCCAGCGCCGCCGCCGGT 4248

RESULT 59
ACL64405/c
ID   ACL64405 standard; DNA; 7874 BP.
XX
AC   ACL64405;
XX
DT   02-JUN-2005 (first entry)
XX
DE   M. xanthus DNA fragment, seq id 868.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS   Myxococcus xanthus.
XX

PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 1; SEQ ID NO 868; 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC   set of 1849 contig and singleton sequences comprising coding sequences,
CC   DNA replication elements, promoters and other regulatory elements from
CC   the genome of the bacterium Myxococcus xanthus. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO.
XX
SQ   Sequence 7874 BP; 1293 A; 2530 C; 2661 G; 1390 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 7874
Score: 1.79e+03   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 14   Gaps: 0

US-09-989-890-238 (1-212) x ACL64405 (1-7874)

Qy   56 SerAlaSerAlaAlaAlaGlyIleAla 64
            |||||||||||||||||||||||||||
Db   1546 AGTGCCAGCGCGGCCGCTGCGGGAATCGCT 1520

RESULT 60
AAS28527/c
ID   AAS28527 standard; DNA; 7880 BP.
XX
AC   AAS28527;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Genomic sequence #367 encoding for novel human respiratory antigen.
XX
KW   Human; respiratory antigen; respiratory disorder; throat disorder;
KW   lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
KW   anti allergic; anti asthmatic; anti inflammatory; anti
KW   respiratory active; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155448-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001333.
XX

PR   31-JAN-2000;   2000US-0179065P.
PR   04-FEB-2000;   2000US-0180628P.
PR   24-FEB-2000;   2000US-0184664P.
PR   02-MAR-2000;   2000US-0186350P.
PR   16-MAR-2000;   2000US-0189874P.
PR   17-MAR-2000;   2000US-0190076P.
PR   18-APR-2000;   2000US-0198123P.
PR   19-MAY-2000;   2000US-0205515P.
PR   07-JUN-2000;   2000US-0209467P.
PR   28-JUN-2000;   2000US-0214486P.
PR   30-JUN-2000;   2000US-0215135P.
PR   07-JUL-2000;   2000US-0216647P.
PR   07-JUL-2000;   2000US-0216880P.
PR   11-JUL-2000;   2000US-0217487P.
PR   11-JUL-2000;   2000US-0217496P.
PR   14-JUL-2000;   2000US-0218290P.
PR   26-JUL-2000;   2000US-0220963P.
PR   26-JUL-2000;   2000US-0220964P.
PR   14-AUG-2000;   2000US-0224518P.
PR   14-AUG-2000;   2000US-0224519P.
PR   14-AUG-2000;   2000US-0225213P.
PR   14-AUG-2000;   2000US-0225214P.
PR   14-AUG-2000;   2000US-0225266P.
PR   14-AUG-2000;   2000US-0225267P.
PR   14-AUG-2000;   2000US-0225268P.
PR   14-AUG-2000;   2000US-0225270P.
PR   14-AUG-2000;   2000US-0225447P.
PR   14-AUG-2000;   2000US-0225757P.
PR   14-AUG-2000;   2000US-0225758P.
PR   14-AUG-2000;   2000US-0225759P.
PR   18-AUG-2000;   2000US-0226279P.
PR   22-AUG-2000;   2000US-0226681P.
PR   22-AUG-2000;   2000US-0226868P.
PR   22-AUG-2000;   2000US-0227182P.
PR   23-AUG-2000;   2000US-0227009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
XX

```
PR   02-OCT-2000;  2000US-0237037P.
PR   02-OCT-2000;  2000US-0237038P.
PR   02-OCT-2000;  2000US-0237039P.
PR   13-OCT-2000;  2000US-0239935P.
PR   13-OCT-2000;  2000US-0239937P.
PR   20-OCT-2000;  2000US-0240960P.
PR   20-OCT-2000;  2000US-0241221P.
PR   20-OCT-2000;  2000US-0241785P.
PR   20-OCT-2000;  2000US-0241786P.
PR   20-OCT-2000;  2000US-0241787P.
PR   20-OCT-2000;  2000US-0241808P.
PR   20-OCT-2000;  2000US-0241809P.
PR   20-OCT-2000;  2000US-0241826P.
PR   01-NOV-2000;  2000US-0244617P.
PR   08-NOV-2000;  2000US-0246474P.
PR   08-NOV-2000;  2000US-0246475P.
PR   08-NOV-2000;  2000US-0246476P.
PR   08-NOV-2000;  2000US-0246477P.
PR   08-NOV-2000;  2000US-0246478P.
PR   08-NOV-2000;  2000US-0246523P.
PR   08-NOV-2000;  2000US-0246524P.
PR   08-NOV-2000;  2000US-0246525P.
PR   08-NOV-2000;  2000US-0246526P.
PR   08-NOV-2000;  2000US-0246527P.
PR   08-NOV-2000;  2000US-0246528P.
PR   08-NOV-2000;  2000US-0246532P.
PR   08-NOV-2000;  2000US-0246609P.
PR   08-NOV-2000;  2000US-0246610P.
PR   08-NOV-2000;  2000US-0246611P.
PR   17-NOV-2000;  2000US-0249207P.
PR   17-NOV-2000;  2000US-0249208P.
PR   17-NOV-2000;  2000US-0249209P.
PR   17-NOV-2000;  2000US-0249210P.
PR   17-NOV-2000;  2000US-0249211P.
PR   17-NOV-2000;  2000US-0249212P.
PR   17-NOV-2000;  2000US-0249213P.
PR   17-NOV-2000;  2000US-0249214P.
PR   17-NOV-2000;  2000US-0249215P.
PR   17-NOV-2000;  2000US-0249216P.
PR   17-NOV-2000;  2000US-0249217P.
PR   17-NOV-2000;  2000US-0249218P.
PR   17-NOV-2000;  2000US-0249244P.
PR   17-NOV-2000;  2000US-0249245P.
PR   17-NOV-2000;  2000US-0249264P.
PR   17-NOV-2000;  2000US-0249265P.
PR   17-NOV-2000;  2000US-0249297P.
PR   17-NOV-2000;  2000US-0249299P.
PR   17-NOV-2000;  2000US-0249300P.
PR   01-DEC-2000;  2000US-0250160P.
PR   01-DEC-2000;  2000US-0250391P.
PR   05-DEC-2000;  2000US-0251030P.
PR   05-DEC-2000;  2000US-0251988P.
PR   05-DEC-2000;  2000US-0256719P.
PR   06-DEC-2000;  2000US-0251479P.
PR   08-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
PR   08-DEC-2000;  2000US-0251989P.
PR   08-DEC-2000;  2000US-0251990P.
PR   11-DEC-2000;  2000US-0254097P.
PR   05-JAN-2001;  2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476224/51.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the respiratory system including respiratory cancers
PT   and also for testing and detection e.g. diagnosis.
```

```
XX
PS   Disclosure; SBD ID No 961; 546pp; English.
XX
CC   The present invention relates to the isolation of novel human respiratory
CC   antigens (AAU17685-AAU17975), and cDNA and genomic sequences encoding for
CC   these polypeptides. The sequences of the invention are useful for
CC   preventing, treating and/or prognosing disorders related to the
CC   respiratory system including throat disorders (e.g. vocal cord paralysis,
CC   tonsilitis, and laryngitis; lung disorders e.g. pneumonia, allergic
CC   disorders e.g. asthma, pleurisy, cystic fibrosis, emphysema, nose
CC   disorders and cancers of the respiratory tissues e.g. lung cancer. The
CC   polynucleotide sequences of the invention are useful in gene therapy and
CC   antisense therapy. AAS28161-AAS28764 represent genomic sequences encoding
CC   for novel human respiratory antigens. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7880 BP; 1588 A; 2332 C; 2092 G; 1868 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                  1.79e+03        Length:          7880
Score:                      9.00            Matches:            9
Percent Similarity:         100.0%          Conservative:       0
Best Local Similarity:      100.0%          Mismatches:         0
Query Match:                4.2%            Indels:             0
DB:                         4               Gaps:               0

US-09-989-890-238 (1-212) x AAS28527 (1-7880)

Qy         25 ProSerLeuAlaLeuLeuSerArg 33
              ||||||||||||||||||||||||
Db       4716 CCGAGCTCCCTGGCTCTACTATCCAGG 4690


RESULT 61
AAS28526/c
ID   AAS28526 standard; DNA; 7880 BP.
XX
AC   AAS28526;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Genomic sequence #366 encoding for novel human respiratory antigen.
XX
KW   Human; respiratory antigen; respiratory disorder; throat disorder;
KW   lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
KW   anti allergic; anti asthmatic; anti inflammatory; olfactory;
KW   respiratory active; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155448-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001333.
XX
PR   31-JAN-2000;  2000US-0179065P.
PR   04-FEB-2000;  2000US-0180628P.
PR   24-FEB-2000;  2000US-0184664P.
PR   02-MAR-2000;  2000US-0186350P.
PR   16-MAR-2000;  2000US-0189874P.
PR   17-MAR-2000;  2000US-0190076P.
PR   18-APR-2000;  2000US-0198123P.
PR   19-MAY-2000;  2000US-0205515P.
PR   07-JUN-2000;  2000US-0209467P.
PR   28-JUN-2000;  2000US-0214886P.
PR   30-JUN-2000;  2000US-0215135P.
PR   07-JUL-2000;  2000US-0216647P.
PR   07-JUL-2000;  2000US-0216880P.
PR   11-JUL-2000;  2000US-0217487P.
PR   11-JUL-2000;  2000US-0217496P.
PR   14-JUL-2000;  2000US-0218290P.
```

```
PR   26-JUL-2000;   2000US-0209963P.          PR   08-NOV-2000;   2000US-0246475P.
PR   26-JUL-2000;   2000US-0209964P.          PR   08-NOV-2000;   2000US-0246476P.
PR   14-AUG-2000;   2000US-0224518P.          PR   08-NOV-2000;   2000US-0246477P.
PR   14-AUG-2000;   2000US-0224519P.          PR   08-NOV-2000;   2000US-0246478P.
PR   14-AUG-2000;   2000US-0225213P.          PR   08-NOV-2000;   2000US-0246523P.
PR   14-AUG-2000;   2000US-0225214P.          PR   08-NOV-2000;   2000US-0246524P.
PR   14-AUG-2000;   2000US-0225266P.          PR   08-NOV-2000;   2000US-0246525P.
PR   14-AUG-2000;   2000US-0225267P.          PR   08-NOV-2000;   2000US-0246526P.
PR   14-AUG-2000;   2000US-0225268P.          PR   08-NOV-2000;   2000US-0246527P.
PR   14-AUG-2000;   2000US-0225270P.          PR   08-NOV-2000;   2000US-0246528P.
PR   14-AUG-2000;   2000US-0225447P.          PR   08-NOV-2000;   2000US-0246532P.
PR   14-AUG-2000;   2000US-0225757P.          PR   08-NOV-2000;   2000US-0246609P.
PR   14-AUG-2000;   2000US-0225758P.          PR   08-NOV-2000;   2000US-0246610P.
PR   18-AUG-2000;   2000US-0225759P.          PR   08-NOV-2000;   2000US-0246611P.
PR   22-AUG-2000;   2000US-0226279P.          PR   08-NOV-2000;   2000US-0246613P.
PR   22-AUG-2000;   2000US-0226681P.          PR   17-NOV-2000;   2000US-0249207P.
PR   22-AUG-2000;   2000US-0226868P.          PR   17-NOV-2000;   2000US-0249208P.
PR   23-AUG-2000;   2000US-0227182P.          PR   17-NOV-2000;   2000US-0249209P.
PR   30-AUG-2000;   2000US-0227009P.          PR   17-NOV-2000;   2000US-0249210P.
PR   01-SEP-2000;   2000US-0228924P.          PR   17-NOV-2000;   2000US-0249211P.
PR   01-SEP-2000;   2000US-0229287P.          PR   17-NOV-2000;   2000US-0249212P.
PR   01-SEP-2000;   2000US-0229343P.          PR   17-NOV-2000;   2000US-0249213P.
PR   05-SEP-2000;   2000US-0229344P.          PR   17-NOV-2000;   2000US-0249214P.
PR   05-SEP-2000;   2000US-0229345P.          PR   17-NOV-2000;   2000US-0249215P.
PR   06-SEP-2000;   2000US-0229509P.          PR   17-NOV-2000;   2000US-0249216P.
PR   06-SEP-2000;   2000US-0229513P.          PR   17-NOV-2000;   2000US-0249217P.
PR   08-SEP-2000;   2000US-0230437P.          PR   17-NOV-2000;   2000US-0249218P.
PR   08-SEP-2000;   2000US-0230438P.          PR   17-NOV-2000;   2000US-0249244P.
PR   08-SEP-2000;   2000US-0231242P.          PR   17-NOV-2000;   2000US-0249245P.
PR   08-SEP-2000;   2000US-0231243P.          PR   17-NOV-2000;   2000US-0249264P.
PR   08-SEP-2000;   2000US-0231244P.          PR   17-NOV-2000;   2000US-0249265P.
PR   12-SEP-2000;   2000US-0231413P.          PR   17-NOV-2000;   2000US-0249297P.
PR   14-SEP-2000;   2000US-0231414P.          PR   17-NOV-2000;   2000US-0249299P.
PR   14-SEP-2000;   2000US-0232080P.          PR   17-NOV-2000;   2000US-0249300P.
PR   14-SEP-2000;   2000US-0232081P.          PR   01-DEC-2000;   2000US-0250160P.
PR   14-SEP-2000;   2000US-0232397P.          PR   01-DEC-2000;   2000US-0250391P.
PR   14-SEP-2000;   2000US-0232398P.          PR   05-DEC-2000;   2000US-0251030P.
PR   14-SEP-2000;   2000US-0232399P.          PR   05-DEC-2000;   2000US-0251988P.
PR   14-SEP-2000;   2000US-0232400P.          PR   05-DEC-2000;   2000US-0256719P.
PR   21-SEP-2000;   2000US-0232401P.          PR   06-DEC-2000;   2000US-0251479P.
PR   21-SEP-2000;   2000US-0233063P.          PR   08-DEC-2000;   2000US-0251856P.
PR   25-SEP-2000;   2000US-0233064P.          PR   08-DEC-2000;   2000US-0251868P.
PR   26-SEP-2000;   2000US-0233065P.          PR   08-DEC-2000;   2000US-0251869P.
PR   27-SEP-2000;   2000US-0234223P.          PR   08-DEC-2000;   2000US-0251989P.
PR   29-SEP-2000;   2000US-0234274P.          PR   08-DEC-2000;   2000US-0251990P.
PR   29-SEP-2000;   2000US-0234997P.          PR   11-DEC-2000;   2000US-0254097P.
PR   29-SEP-2000;   2000US-0234998P.          PR   05-JAN-2001;   2001US-0259678P.
PR   29-SEP-2000;   2000US-0235484P.          XX
PR   02-OCT-2000;   2000US-0235834P.          PA   (HUMA-) HUMAN GENOME SCI INC.
PR   02-OCT-2000;   2000US-0235836P.          XX
PR   13-OCT-2000;   2000US-0236327P.          PI   Rosen CA,  Barash SC,  Ruben SM;
PR   20-OCT-2000;   2000US-0236367P.          XX
PR   20-OCT-2000;   2000US-0236369P.          DR   WPI; 2001-476224/51.
PR   20-OCT-2000;   2000US-0236370P.          XX
PR   20-OCT-2000;   2000US-0236802P.          PT   Isolated polypeptide for treating, preventing and/ or prognosing
PR   20-OCT-2000;   2000US-0237037P.          PT   disorders related to the respiratory system including respiratory cancers
PR   20-OCT-2000;   2000US-0237038P.          PT   and also for testing and detection e.g. diagnosis.
PR   20-OCT-2000;   2000US-0237039P.          XX
PR   02-OCT-2000;   2000US-0237040P.          PS   Disclosure; SEQ ID No 960; 546pp; English.
PR   20-OCT-2000;   2000US-0239935P.          XX
PR   20-OCT-2000;   2000US-0239937P.          CC   The present invention relates to the isolation of novel human respiratory
PR   20-OCT-2000;   2000US-0240960P.          CC   antigens (AAU17685-AAU17975), and cDNA and genomic sequences encoding for
PR   20-OCT-2000;   2000US-0241221P.          CC   these polypeptides. The sequences of the invention are useful for
PR   20-OCT-2000;   2000US-0241785P.          CC   preventing, treating and/or prognosing disorders related to the
PR   20-OCT-2000;   2000US-0241786P.          CC   respiratory system including throat disorders (e.g. vocal cord paralysis,
PR   20-OCT-2000;   2000US-0241787P.          CC   tonsillitis, and laryngitis) lung disorders e.g. pneumonia, allergic
PR   20-OCT-2000;   2000US-0241808P.          CC   disorders e.g. asthma, pleurisy, cystic fibrosis, emphysema, nose
PR   20-OCT-2000;   2000US-0241809P.          CC   disorders and cancers of the respiratory tissues e.g. lung cancer. The
PR   20-OCT-2000;   2000US-0241826P.          CC   polynucleotide sequences of the invention are useful in gene therapy and
PR   01-NOV-2000;   2000US-0244617P.          CC   antisense therapy. AAS28161-AAS28764 represent genomic sequences encoding
PR   08-NOV-2000;   2000US-0246474P.          CC   for novel human respiratory antigens. Note: The sequence data for this
                                              CC   patent did not form part of the printed specification, but was obtained
                                              CC   in electronic format directly from WIPO at
```

```
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7880 BP; 1588 A; 2333 C; 2091 G; 1868 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.79e+03    Length:              7880
Score:                  9.00        Matches:             9
Percent Similarity:     100.0%      Conservative:        0
Best Local Similarity:  100.0%      Mismatches:          0
Query Match:            4.2%        Indels:              0
DB:                                 Gaps:                0

US-09-989-890-238 (1-212) x AAS28526 (1-7880)

Qy        25 ProSerLeuAlaLeuLeuSerArg 33
             ||||||||||||||||||||||||
Db      4716 CCGAGCTCCCTGGCTCTACTATCCAGG 4690

RESULT 62
ADG41723/c
ID   ADG41723 standard; DNA; 7880 BP.
XX
AC   ADG41723;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Human respiratory system associated genomic DNA seq id 961.
XX
KW   antiinflammatory; antiallergic; antiasthmatic; cytostatic; gene therapy;
KW   respiratory system antigen;
KW   human respiratory system associated polynucleotide;
KW   respiratory system disorder; throat disorder; vocal cord paralysis;
KW   tonsillitis; laryngitis; lung disorder; pneumonia; allergic disorder;
KW   asthma; eosinophilic pneumonia; pleurisy; cystic fibrosis; emphysema;
KW   histiocytosis; sarcoidosis; nose disorder; rhinitis; sinusitis; neoplasm;
KW   cancer; respiratory tissue cancer; throat cancer; lung cancer;
KW   cancer of the nose; gene therapy; chromosome identification; forensic;
KW   human respiratory system associated protein; ds; human.
XX
OS   Homo sapiens.
XX
PN   US2003215893-A1.
XX
PD   20-NOV-2003.
XX
PF   07-AUG-2002; 2002US-00212872.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216477P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   28-SEP-2000; 2000US-0235935P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236677P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
```

```
CC    techniques, for chromosome identification, identifying individuals from

Alignment Scores:
Pred. No.:                1.79e+03        Length:         7880
Score:                    9.00            Matches:        9
Percent Similarity:       100.0%          Conservative:   0
Best Local Similarity:    100.0%          Mismatches:     0
Query Match:              4.2%            Indels:         0
DB:                                        Gaps:           0

US-09-989-890-238 (1-212) x ADG41723 (1-7880)

Qy        25 ProSerSerLeuAlaLeuLeuSerArg 33
              |||||||||||||||||||||||||||
Db      4716 CCGAGCTCCCTGGCTCTACTATCCAGG 4690

RESULT 63
ADG41722/c
ID   ADG41722 standard; DNA; 7880 BP.
XX
AC   ADG41722;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Human respiratory system associated genomic DNA seq id 960.
XX
KW   antiinflammatory; antiallergic; antiasthmatic; cytostatic; gene therapy;
KW   respiratory system antigen;
KW   human respiratory system associated polynucleotide;
KW   respiratory system disorder; throat disorder; vocal cord paralysis;
KW   tonsillitis; laryngitis; lung disorder; pneumonia; allergic disorder;
KW   asthma; eosinophilic pneumonia; pleurisy; cystic fibrosis; emphysema;
KW   histiocytosis; sarcoidosis; nose disorder; rhinitis; sinusitis; neoplasm;
KW   cancer; respiratory tissue cancer; throat cancer; lung cancer;
KW   cancer of the nose; gene therapy; chromosome identification; forensic;
KW   human respiratory system associated protein; ds; human.
XX
OS   Homo sapiens.
XX
PN   US2003215893-A1.
XX
PD   20-NOV-2003.
XX
PF   07-AUG-2002; 2002US-00212872.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225135P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
```

```
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0254479P.
PR   08-DEC-2000; 2000US-0258556P.
PR   08-DEC-2000; 2000US-0258668P.
PR   08-DEC-2000; 2000US-0258686P.
PR   08-DEC-2000; 2000US-0259989P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0250990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2000US-0259678P.
PR   17-JAN-2001; 2001US-00764860.
PR   14-FEB-2002; 2002US-00074095.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-902033/82.
XX
PT   Novel respiratory system antigen and polynucleotides encoding the
PT   polypeptides, useful for treating diagnosing, treating or preventing
PT   tonsillitis, pneumonia, asthma and cystic fibrosis, emphysema, throat
PT   cancer.
XX
PS   Disclosure; SEQ ID NO 961; 236pp; English.
XX
CC   The invention describes an isolated polypeptide (I) comprising an amino
CC   acid sequence that is at least 90% identical to polypeptide fragment of
CC   any one of 299 respiratory system antigen sequences (PS) and having
CC   biological activity, polypeptide domain or epitope of PS, full-length
CC   protein of PS, or variant, allelic variant or species homolog of PS. (I)
CC   or a polynucleotide (II) encoding (I) is also useful for diagnosing a
CC   pathological condition or a susceptibility to a pathological condition in
CC   a subject which involves determining the presence or absence of mutation
CC   in (II) or determining the presence or amount of expression of (I) in a
CC   biological sample and diagnosing a pathological condition based on the
CC   result. The human respiratory system associated polynucleotides, the
CC   polypeptides encoded by them, and antibodies that immunospecifically bind
CC   these polypeptides are useful in diagnosis, treatment, prevention and/or
CC   prognosis of disorders of respiratory system such as throat disorders
CC   (e.g., vocal cord paralysis, tonsillitis, and laryngitis), lung disorders
CC   (e.g., pneumonia), allergic disorders, (e.g., asthma and eosinophilic
CC   pneumonia), pleurisy, cystic fibrosis, emphysema, histiocytosis,
CC   sarcoidosis, nose disorders (rhinitis and sinusitis), neoplasms and/or
CC   cancers of respiratory tissues (e.g., throat cancer, lung cancer, and
CC   cancer of the nose). The polynucleotides are useful in gene therapy
```

PR   14-AUG-2000;   2000US-0225758P.
PR   14-AUG-2000;   2000US-0225759P.
PR   18-AUG-2000;   2000US-0226279P.
PR   22-AUG-2000;   2000US-0226681P.
PR   22-AUG-2000;   2000US-0226868P.
PR   22-AUG-2000;   2000US-0227182P.
PR   23-AUG-2000;   2000US-0227009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   28-SEP-2000;   2000US-0235935P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0241786P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0244474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.

PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
PR   17-JAN-2001;   2001US-00764860.
PR   14-FEB-2002;   2002US-00074095.

XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Ruben SM,   Barash SC;
XX
DR   WPI; 2003-902033/82.
XX
PT   Novel respiratory system antigen and polynucleotides encoding the
PT   polypeptides, useful for treating diagnosing, treating or preventing
PT   tonsillitis, pneumonia, asthma and cystic fibrosis, emphysema, throat
PT   cancer.
XX
PS   Disclosure; SEQ ID NO 960; 236pp; English.
XX
CC   The invention describes an isolated polypeptide (I) comprising an amino
CC   acid sequence that is at least 90% identical to polypeptide fragment of
CC   any one of 299 respiratory system antigen sequences (PS) and having
CC   biological activity, polypeptide domain or epitope of PS, full-length
CC   protein of PS, or variant, allelic variant or species homolog of PS. (I)
CC   or a polynucleotide (II) encoding (I) is also useful for diagnosing a
CC   pathological condition or a susceptibility to a pathological condition in
CC   a subject which involves determining the presence or absence of mutation
CC   in (II) or determining the presence or amount of expression of (I) in a
CC   biological sample and diagnosing a pathological condition based on the
CC   result. The human respiratory system associated polynucleotides, the
CC   polypeptides encoded by them, and antibodies that immunospecifically bind
CC   these polypeptides are useful in diagnosis, treatment, prevention and/or
CC   prognosis of disorders of respiratory system such as throat disorders
CC   (e.g. vocal cord paralysis, tonsillitis, and laryngitis), lung disorders
CC   (e.g., pneumonia), allergic disorders, (e.g., asthma and eosinophilic
CC   pneumonia), pleurisy, cystic fibrosis, emphysema, histiocytosis,
CC   sarcoidosis, nose disorders (rhinitis and sinusitis), neoplasms and/or
CC   cancers of respiratory tissues (e.g., throat cancer, lung cancer, and
CC   cancer of the nose). The polynucleotides are useful in gene therapy
CC   techniques, for chromosome identification, identifying individuals from

```
Alignment Scores:
Pred.No.:                   1.79e+03       Length:              7880
Score:                      9.00           Matches:             9
Percent Similarity:         100.0%         Conservative:        0
Best Local Similarity:      100.0%         Mismatches:          0
Query Match:                4.2%           Indels:              0
DB:                         10             Gaps:                0

US-09-989-890-238 (1-212) x ADG41722 (1-7880)

Qy        25 ProSerSerLeuAlaLeuLeuSerArg 33
             |||||||||||||||||||||||||||
Db      4716 CCGAGCTCCCTGGCTCTACTACCAGG 4690

RESULT 64
ADI97497/c
ID   ADI97497 standard; DNA; 7880 BP.
XX
AC   ADI97497;
XX
DT   04-NOV-2004 (first entry)
XX
DE   Human respiratory system associated polypeptide-related DNA SeqID961.
XX
KW   respiratory system-related polypeptide; antiasthmatic; antibacterial;
KW   antiinflammatory; cytostatic; antianaemic; antianaemic; gene therapy;
KW   pneumonia; lung cancer; cystic fibrosis; asthma; sarcoidosis; rhinitis;
KW   anaemia; leukaemia; inflammation; sinusitis;
KW   chronic obstructive pulmonary disease; infectious disease; human; ds.
XX
OS   Homo sapiens.
XX
PN   US2003077704-A1.
XX
PD   24-APR-2003.
XX
PF   14-FEB-2002; 2002US-00074095.
XX
PR   31-JAN-2000; 2000US-0179065P.       PR   30-AUG-2000; 2000US-0229924P.
PR   04-FEB-2000; 2000US-0180628P.       PR   01-SEP-2000; 2000US-0229287P.
PR   24-FEB-2000; 2000US-0184664P.       PR   01-SEP-2000; 2000US-0229343P.
PR   02-MAR-2000; 2000US-0186350P.       PR   01-SEP-2000; 2000US-0229344P.
PR   16-MAR-2000; 2000US-0189874P.       PR   01-SEP-2000; 2000US-0229345P.
PR   17-MAR-2000; 2000US-0190076P.       PR   05-SEP-2000; 2000US-0229509P.
PR   18-APR-2000; 2000US-0198123P.       PR   05-SEP-2000; 2000US-0229513P.
PR   19-MAY-2000; 2000US-0205515P.       PR   06-SEP-2000; 2000US-0230437P.
PR   07-JUN-2000; 2000US-0209467P.       PR   06-SEP-2000; 2000US-0230438P.
PR   28-JUN-2000; 2000US-0214886P.       PR   08-SEP-2000; 2000US-0231242P.
PR   30-JUN-2000; 2000US-0215135P.       PR   08-SEP-2000; 2000US-0231243P.
PR   07-JUL-2000; 2000US-0216647P.       PR   08-SEP-2000; 2000US-0231244P.
PR   07-JUL-2000; 2000US-0216880P.       PR   08-SEP-2000; 2000US-0231413P.
PR   11-JUL-2000; 2000US-0217487P.       PR   08-SEP-2000; 2000US-0231414P.
PR   11-JUL-2000; 2000US-0217496P.       PR   08-SEP-2000; 2000US-0232080P.
PR   14-JUL-2000; 2000US-0218290P.       PR   08-SEP-2000; 2000US-0232081P.
PR   26-JUL-2000; 2000US-0220963P.       PR   12-SEP-2000; 2000US-0231968P.
PR   26-JUL-2000; 2000US-0220964P.       PR   14-SEP-2000; 2000US-0232397P.
PR   14-AUG-2000; 2000US-0224518P.       PR   14-SEP-2000; 2000US-0232398P.
PR   14-AUG-2000; 2000US-0224519P.       PR   14-SEP-2000; 2000US-0232399P.
PR   14-AUG-2000; 2000US-0225213P.       PR   14-SEP-2000; 2000US-0232400P.
PR   14-AUG-2000; 2000US-0225214P.       PR   14-SEP-2000; 2000US-0232401P.
PR   14-AUG-2000; 2000US-0225266P.       PR   14-SEP-2000; 2000US-0230063P.
PR   14-AUG-2000; 2000US-0225267P.       PR   14-SEP-2000; 2000US-0230064P.
PR   14-AUG-2000; 2000US-0225268P.       PR   14-SEP-2000; 2000US-0230065P.
PR   14-AUG-2000; 2000US-0225270P.       PR   21-SEP-2000; 2000US-0234223P.
PR   14-AUG-2000; 2000US-0225447P.       PR   21-SEP-2000; 2000US-0234274P.
PR   14-AUG-2000; 2000US-0225757P.       PR   25-SEP-2000; 2000US-0234997P.
PR   14-AUG-2000; 2000US-0225758P.       PR   25-SEP-2000; 2000US-0234998P.
PR   14-AUG-2000; 2000US-0225759P.       PR   26-SEP-2000; 2000US-0235484P.
PR   18-AUG-2000; 2000US-0226279P.       PR   27-SEP-2000; 2000US-0235834P.
PR   22-AUG-2000; 2000US-0226681P.       PR   27-SEP-2000; 2000US-0235836P.
PR   22-AUG-2000; 2000US-0226868P.       PR   29-SEP-2000; 2000US-0236327P.
PR   22-AUG-2000; 2000US-0227182P.       PR   29-SEP-2000; 2000US-0236367P.
PR   23-AUG-2000; 2000US-0227009P.       PR   29-SEP-2000; 2000US-0236368P.
                                         PR   29-SEP-2000; 2000US-0236369P.
                                         PR   29-SEP-2000; 2000US-0236370P.
                                         PR   02-OCT-2000; 2000US-0236802P.
                                         PR   02-OCT-2000; 2000US-0237037P.
                                         PR   02-OCT-2000; 2000US-0237038P.
                                         PR   02-OCT-2000; 2000US-0237039P.
                                         PR   02-OCT-2000; 2000US-0237040P.
                                         PR   13-OCT-2000; 2000US-0239935P.
                                         PR   20-OCT-2000; 2000US-0239937P.
                                         PR   20-OCT-2000; 2000US-0240960P.
                                         PR   20-OCT-2000; 2000US-0241221P.
                                         PR   20-OCT-2000; 2000US-0241785P.
                                         PR   20-OCT-2000; 2000US-0241786P.
                                         PR   20-OCT-2000; 2000US-0241787P.
                                         PR   20-OCT-2000; 2000US-0241808P.
                                         PR   20-OCT-2000; 2000US-0241809P.
                                         PR   20-OCT-2000; 2000US-0241826P.
                                         PR   01-NOV-2000; 2000US-0244617P.
                                         PR   08-NOV-2000; 2000US-0246474P.
                                         PR   08-NOV-2000; 2000US-0246475P.
                                         PR   08-NOV-2000; 2000US-0246476P.
                                         PR   08-NOV-2000; 2000US-0246477P.
                                         PR   08-NOV-2000; 2000US-0246478P.
                                         PR   08-NOV-2000; 2000US-0246523P.
                                         PR   08-NOV-2000; 2000US-0246524P.
                                         PR   08-NOV-2000; 2000US-0246525P.
                                         PR   08-NOV-2000; 2000US-0246526P.
                                         PR   08-NOV-2000; 2000US-0246527P.
                                         PR   08-NOV-2000; 2000US-0246528P.
                                         PR   08-NOV-2000; 2000US-0246532P.
                                         PR   08-NOV-2000; 2000US-0246609P.
                                         PR   08-NOV-2000; 2000US-0246610P.
                                         PR   08-NOV-2000; 2000US-0246611P.
                                         PR   08-NOV-2000; 2000US-0246613P.
                                         PR   17-NOV-2000; 2000US-0249207P.
                                         PR   17-NOV-2000; 2000US-0249208P.
                                         PR   17-NOV-2000; 2000US-0249209P.
                                         PR   17-NOV-2000; 2000US-0249210P.
```

```
PR    17-NOV-2000;   2000US-0249211P.
PR    17-NOV-2000;   2000US-0249212P.
PR    17-NOV-2000;   2000US-0249213P.
PR    17-NOV-2000;   2000US-0249214P.
PR    17-NOV-2000;   2000US-0249215P.
PR    17-NOV-2000;   2000US-0249216P.
PR    17-NOV-2000;   2000US-0249217P.
PR    17-NOV-2000;   2000US-0249218P.
PR    17-NOV-2000;   2000US-0249244P.
PR    17-NOV-2000;   2000US-0249245P.
PR    17-NOV-2000;   2000US-0249264P.
PR    17-NOV-2000;   2000US-0249265P.
PR    17-NOV-2000;   2000US-0249297P.
PR    17-NOV-2000;   2000US-0249299P.
PR    17-NOV-2000;   2000US-0249300P.
PR    01-DEC-2000;   2000US-0250160P.
PR    01-DEC-2000;   2000US-0250391P.
PR    05-DEC-2000;   2000US-0251030P.
PR    05-DEC-2000;   2000US-0251988P.
PR    05-DEC-2000;   2000US-0256719P.
PR    06-DEC-2000;   2000US-0251479P.
PR    08-DEC-2000;   2000US-0251856P.
PR    08-DEC-2000;   2000US-0251868P.
PR    08-DEC-2000;   2000US-0251869P.
PR    08-DEC-2000;   2000US-0251989P.
PR    08-DEC-2000;   2000US-0251990P.
PR    11-DEC-2000;   2000US-0254097P.
PR    05-JAN-2001;   2001US-0259678P.
PR    17-JAN-2001;   2001US-00764860.
XX
PA    (HUMA-)  HUMAN GENOME SCI INC.
XX
PI    Rosen CA,  Ruben SM,  Barash SC;
XX
DR    WPI; 2003-765403/72.
XX
PT    New human respiratory system-related polypeptide and genes, useful for
PT    treating, preventing or diagnosing e.g. pneumonia, lung cancer, cystic
PT    fibrosis, asthma, sarcoidosis, rhinitis, leukemia, inflammations or
PT    sinusitis.
XX
PS    Disclosure; SEQ ID NO 961; 202pp; English.
XX
CC    This invention is related to a novel isolated polypeptide, which
CC    comprises a human respiratory system-related polypeptide, and the DNA
CC    sequence which encodes it. The invention may be useful for the
CC    development of compounds with an antiasthmatic, antibacterial,
CC    antiinflammatory, cytostatic, antianaemic or antiallergic activity. In
CC    addition, the sequences disclosed may be useful for gene therapy. The
CC    polypeptide or polynucleotide is useful for treating, preventing or
CC    ameliorating a medical condition, for example pneumonia, lung cancer,
CC    cystic fibrosis, asthma, sarcoidosis, rhinitis, anaemia, leukaemia,
CC    inflammations, sinusitis, chronic obstructive pulmonary disease or
CC    infectious diseases. The polypeptide or polynucleotide is also useful for
CC    diagnosing any of these diseases or a susceptibility to the disease. The
CC    present sequence is that of a human DNA sequence which is related to a
CC    human respiratory system associated gene of the invention.
XX
SQ    Sequence 7880 BP; 1588 A; 2332 C; 2092 G; 1868 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                            1.79e+03        Length:         7880
Score:                                9.00            Matches:        9
Percent Similarity:                   100.0%          Conservative:   9
Best Local Similarity:                100.0%          Mismatches:     0
Query Match:                          4.2%            Indels:         0
DB:                                   11              Gaps:           0

US-09-989-890-238 (1-212)  x ADI97497 (1-7880)

Qy         25  ProSerSerLeuAlaLeuLeuSerArg  33
               |||||||||||||||||||||||||||
Db       4716  CCGAGCTCCCTGGCTCTACTATCCAGG  4690
```

```
RESULT 65
AD197496/c
ID    AD197496 standard; DNA; 7880 BP.
XX
AC    AD197496;
XX
DT    04-NOV-2004  (first entry)
XX
DE    Human respiratory system associated polypeptide-related DNA SeqID960.
XX
KW    respiratory system-related polypeptide; antiasthmatic; antibacterial;
KW    antiinflammatory; cytostatic; antianaemic; antiallergic; gene therapy;
KW    pneumonia; lung cancer; cystic fibrosis; asthma; sarcoidosis; rhinitis;
KW    anaemia; leukaemia; inflammation; sinusitis;
KW    chronic obstructive pulmonary disease; infectious disease; human; ds.
XX
OS    Homo sapiens.
XX
PN    US2003077704-A1.
XX
PD    24-APR-2003.
XX
PF    14-FEB-2002; 2002US-00074095.
XX
PR    31-JAN-2000;  2000US-0179065P.
PR    04-FEB-2000;  2000US-0180628P.
PR    24-FEB-2000;  2000US-0184664P.
PR    02-MAR-2000;  2000US-0186350P.
PR    16-MAR-2000;  2000US-0189874P.
PR    17-MAR-2000;  2000US-0190076P.
PR    18-APR-2000;  2000US-0198123P.
PR    19-MAY-2000;  2000US-0205515P.
PR    07-JUN-2000;  2000US-0209467P.
PR    28-JUN-2000;  2000US-0214886P.
PR    30-JUN-2000;  2000US-0215135P.
PR    07-JUL-2000;  2000US-0216647P.
PR    07-JUL-2000;  2000US-0216880P.
PR    11-JUL-2000;  2000US-0217487P.
PR    11-JUL-2000;  2000US-0217496P.
PR    14-JUL-2000;  2000US-0218290P.
PR    26-JUL-2000;  2000US-0220963P.
PR    26-JUL-2000;  2000US-0220964P.
PR    14-AUG-2000;  2000US-0224518P.
PR    14-AUG-2000;  2000US-0224519P.
PR    14-AUG-2000;  2000US-0225213P.
PR    14-AUG-2000;  2000US-0225214P.
PR    14-AUG-2000;  2000US-0225266P.
PR    14-AUG-2000;  2000US-0225267P.
PR    14-AUG-2000;  2000US-0225268P.
PR    14-AUG-2000;  2000US-0225270P.
PR    14-AUG-2000;  2000US-0225447P.
PR    14-AUG-2000;  2000US-0225757P.
PR    14-AUG-2000;  2000US-0225758P.
PR    14-AUG-2000;  2000US-0225759P.
PR    18-AUG-2000;  2000US-0226279P.
PR    22-AUG-2000;  2000US-0226681P.
PR    22-AUG-2000;  2000US-0226868P.
PR    22-AUG-2000;  2000US-0227182P.
PR    23-AUG-2000;  2000US-0227009P.
PR    30-AUG-2000;  2000US-0228924P.
PR    01-SEP-2000;  2000US-0229287P.
PR    01-SEP-2000;  2000US-0229343P.
PR    01-SEP-2000;  2000US-0229344P.
PR    01-SEP-2000;  2000US-0229345P.
PR    05-SEP-2000;  2000US-0229509P.
PR    05-SEP-2000;  2000US-0229513P.
PR    06-SEP-2000;  2000US-0230437P.
PR    06-SEP-2000;  2000US-0230438P.
PR    08-SEP-2000;  2000US-0231242P.
PR    08-SEP-2000;  2000US-0231243P.
PR    08-SEP-2000;  2000US-0231244P.
PR    08-SEP-2000;  2000US-0231413P.
```

PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.

PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
PR   17-JAN-2001;   2001US-00764860.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-765403/72.
XX
PT   New human respiratory system-related polypeptide and genes, useful for
PT   treating, preventing or diagnosing e.g. pneumonia, lung cancer, cystic
PT   fibrosis, asthma, sarcoidosis, rhinitis, leukemia, inflammations or
PT   sinusitis.
XX
PS   Disclosure; SEQ ID NO 960; 202pp; English.
XX
CC   This invention is related to a novel isolated polypeptide, which
CC   comprises a human respiratory system-related polypeptide, and the DNA
CC   sequence which encodes it. The invention may be useful for the
CC   development of compounds with an antiasthmatic, antibacterial,
CC   antiinflammatory, cytostatic, antianaemic or antiallergic activity. In
CC   addition, the sequences disclosed may be useful for gene therapy. The
CC   polypeptide or polynucleotide is useful for treating, preventing or
CC   ameliorating a medical condition, for example pneumonia, lung cancer,
CC   cystic fibrosis, asthma, sarcoidosis, rhinitis, anaemia, leukaemia,
CC   inflammations, sinusitis, chronic obstructive pulmonary disease or
CC   infectious diseases. The polypeptide or polynucleotide is also useful for
CC   diagnosing any of these diseases or a susceptibility to the disease. The
CC   present sequence is that of a human DNA sequence which is related to a
CC   human respiratory system associated gene of the invention.
XX
SQ   Sequence 7880 BP; 1588 A; 2333 C; 2091 G; 1868 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   1.79e+03   Length: 7880
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 11   Gaps: 0

US-09-989-890-238 (1-212) x ADI97496 (1-7880)

Qy   25 ProSerLeuAlaLeuLeuSerArg 33
                  |||||||||||||||||||||||
Db   4716 CCGAGCTCCCTGGCTCACTATCCAGG 4690

RESULT 66
ACL64503/c
ID   ACL64503 standard; DNA; 9198 BP.
XX
AC   ACL64503;
XX
DT   02-JUN-2005   (first entry)
XX
DE   M. xanthus DNA fragment, seq id 966.
XX
KW   Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX

OS   Myxococcus xanthus.
XX
PN   US6833447-B1.
XX
PD   21-DEC-2004.
XX
PF   10-JUL-2001; 2001US-00902540.
XX
PR   10-JUL-2000; 2000US-0217883P.
XX
PA   (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI   Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR   WPI; 2005-028716/03.
XX
PT   New substantially purified Myxococcus xanthus nucleic acid molecule
PT   encoding a nitrite reductase, useful for determining gene expression,
PT   identifying mutations in a gene of interest, and for constructing
PT   mutations in a gene of interest.
XX
PS   Example 1; SEQ ID NO 966;· 25pp; English.
XX
CC   The invention relates to a substantially purified nucleic acid molecule
CC   encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC   recombinant DNA construct for expression of a nitrite reductase gene in a
CC   plant cell, and a plant cell comprising the recombinant DNA construct.
CC   The nucleic acid is useful for determining gene expression, identifying
CC   mutations in a gene of interest, and for constructing mutations in a gene
CC   of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC   set of 1849 contig and singleton sequences comprising coding sequences,
CC   DNA replication elements, promoters and other regulatory elements from
CC   the genome of the bacterium Myxococcus xanthus. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO
XX
SQ   Sequence 9198 BP; 1247 A; 3175 C; 3389 G; 1387 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 9198
ABX10394/c
Score: 2.06e+03   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 14

US-09-989-890-238 (1-212) x ACL64503 (1-9198)

Qy   50 SerGlyAlaGlyLeuProSerAlaSer 58
           ||||||||||||||||||||||||||
Db   1342 TCAGGTGCTGGCCTTCCTTCAGCGAGT 1316

RESULT 67
ID   ABX10394 standard; DNA; 9771 BP.
XX
AC   ABX10394;
XX
DT   28-JAN-2003 (first entry)
XX
DE   DNA encoding protein differentially regulated in prostate cancer #63.
XX
KW   Prostate cancer; gene expression; differential regulation;
KW   molecular marker; drug target; cancer detection; cancer diagnosis;
KW   cancer staging; cancer grading; cancer assessing; cancer monitoring;
KW   gene; db.
XX
OS   Homo sapiens.
XX
PN   WO200281638-A2.
XX
PD   17-OCT-2002.
XX

PF   08-APR-2002; 2002WO-US010824.
XX
PR   06-APR-2001; 2001US-0281731P.
PR   06-APR-2001; 2001US-0281732P.
XX
PA   (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI   Sun Z, Jay G;
XX
DR   WPI; 2003-058520/05.
DR   P-PSDB; ABU07489.
XX
PT   Novel genes which are differentially regulated in prostate cancer, useful
PT   for diagnosing prostate cancer in prostate tissue sample and assessing
PT   therapeutic or preventive intervention in prostate cancer patients.
XX
PS   Claim 1; Page 140-144; 416pp; English.
XX
CC   The invention describes genes (I) which are differentially regulated in
CC   prostate cancer. (I) is useful for diagnosing a prostate cancer in a
CC   sample comprising prostate tissue, which involves determining the number
CC   of target genes which are differentially-regulated in the sample, where
CC   the number is indicative of the probability that the sample comprises
CC   prostate cancer. (I) is useful for assessing a therapeutic or preventive
CC   intervention in a subject having a prostate cancer, which involves
CC   determining the expression levels in a sample comprising prostate tissue
CC   of target genes which are differentially-regulated in prostate cancer.
CC   Preferably, the expression levels of at least 10 genes are determined.
CC   (I) is also useful for identifying agents that modulate a biological
CC   activity of a polypeptide differentially-regulated in prostate cancer
CC   cells, which involves contacting a polypeptide differentially-regulated
CC   in prostate cancer cells with a test agent under conditions effective for
CC   the test agent to modulate a biological activity of the polypeptide, and
CC   determining whether the test agent modulates the biological activity. (I)
CC   is useful as molecular markers, as drug targets, and for detecting,
CC   diagnosing, staging, grading, assessing, monitoring, prognosticating,
CC   preventing or treating, determining predisposition to diseases and
CC   conditions especially relating to prostate cancer. (I) and its expression
CC   products are used in the diagnostic test to assay for presence of cancer
CC   e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
CC   blood etc. (I) is useful for assessing cancer e.g., to determine the type
CC   of cancer, its stage of development, the nature of genetic defect, etc.
CC   The polypeptide encoded by (I) can be used as target for therapy or drug
CC   discovery.(I) can also be used for expressing the polypeptide and thus
CC   for searching specific binding partners of the polypeptide. (I) is useful
CC   in therapeutic applications to treat prostate cancer. The identification
CC   of specific genes, and groups of genes, expressed in pathways
CC   physiologically relevant to prostate cancer permits the definition of
CC   functional and disease pathways and the delineation of targets in these
CC   pathways which are useful in diagnostic, therapeutic, and clinical
CC   applications. This sequence encodes a protein differentially regulated in
CC   prostate cancer
XX
SQ   Sequence 9771 BP; 3514 A; 2357 C; 2283 G; 1617 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 9771
Score: 2.18e+03   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 8

US-09-989-890-238 (1-212) x ABX10394 (1-9771)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||
Db   4936 GCTCTCTTTCACGCCCACTTTCCCCA 4910

RESULT 68
ID   ADI02721/c
ID   ADI02721 standard; cDNA; 10432 BP.
XX

```
AC   ADI02721;
XX
DT   22-APR-2004  (first entry)
XX
DE   Human cDNA differentially expressed in the vascular endothelium #262.
XX
KW   88; vascular endothelium; vascular disorder; atherosclerosis;
KW   haemangioma; haemangioendothelioma; oedema; diabetic retinopathy; wart;
KW   pyogenic granuloma; Kaposi's sarcoma; scar keloid; allergic oedema;
KW   neoplasm; psoriasis; ulcer; follicular cyst; endometriosis;
KW   peritoneal sclerosis; obesity; human.
XX
OS   Homo sapiens.
XX
PN   US2003166903-A1.
XX
PD   04-SEP-2003.
XX
PF   25-APR-2002; 2002US-00133013.
XX
PR   27-APR-2001; 2001US-0287067P.
XX
PA   (ASTR/) ASTROMOFF A.
PA   (BAND/) BANDMAN O.
PA   (COCK/) COCKS B G.
XX
PI   Astromoff A, Bandman O, Cocks BG;
XX
DR   WPI; 2003-898115/82.
XX
PT   New combination comprising cDNAs that are differentially expressed in
PT   vascular endothelium, useful for preparing a composition for diagnosing
PT   or treating vascular disorder, e.g., atherosclerosis, psoriasis, ulcers
PT   or obesity.
XX
PS   Claim 1; SEQ ID NO 262; 28pp; English.
XX
CC   The invention relates to a new combination which comprises cDNAs or their
CC   complements not given in the specification that are differentially
CC   expressed in vascular endothelium. The combination of cDNAs is useful for
CC   preparing a composition for diagnosing or treating vascular disorder,
CC   comprising atherosclerosis, haemangioma, haemangioendothelioma, oedema,
CC   diabetic retinopathy, warts, pyogenic granulomas, Kaposi's sarcoma, scar
CC   keloids, allergic oedema, neoplasms, psoriasis, ulcers, follicular cysts,
CC   endometriosis, peritoneal sclerosis or obesity. The present sequence
CC   represents a cDNA differentially expressed in the vascular endothelium.
XX
SQ   Sequence 10432 BP; 3683 A; 2503 C; 2450 G; 1796 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               2.32e+03        Length:          10432
Score:                   9.00            Matches:         9
Percent Similarity:      100.0%          Conservative:    0
Best Local Similarity:   100.0%          Mismatches:      0
Query Match:             4.2%            Indels:          0
DB:                      10              Gaps:            0

US-09-989-890-238 (1-212) x ADI02721 (1-10432)

Qy      29 AlaLeuSerArgProLeuSerPro 37
           |||||||||||||||||||||||||
Db    5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 69
ADJ56543/c
ID   ADJ56543 standard; cDNA; 10432 BP.
XX
AC   ADJ56543;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Murine cDNA differentially expressed in MYCN activated cells SeqID 349.
XX
```

```
KW   mouse; murine; differential expression; transactivator; proto-oncogene;
KW   neuroblastoma; small cell lung cancer; cytostatic; gene therapy; 88;
KW   MYCN activated cell.
XX
OS   Mus musculus.
XX
PN   US2003119009-A1.
XX
PD   26-JUN-2003.
XX
PF   25-FEB-2002; 2002US-00084817.
XX
PR   23-FEB-2001; 2001US-0270784P.
XX
PA   (STUA/) STUART S G.
PA   (NUCH/) NUCHTERN J G.
PA   (PLON/) PLON S E.
PA   (SHOH/) SHOHET J M.
XX
PI   Stuart SG, Nuchtern JG, Plon SE, Shohet JM;
XX
DR   WPI; 2003-635698/60.
XX
PT   New genes regulated by MYCN activation, useful in gene therapy,
PT   particularly for treating a subject with e.g. neuroblastoma or other
PT   cancers, or for diagnosing, staging or monitoring the treatment of the
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 349; 27pp; English.
XX
CC   This invention relates to novel isolated cDNAs that are differentially
CC   expressed in MYCN activated cells. Specifically, it refers to
CC   polynucleotide sequences that exhibit differential expression patterns in
CC   cells activated by the transactivator MYCN, where MYCN is a proto-
CC   oncogene that is amplified in neuroblastoma cells and is common in small
CC   cell lung cancers. The present invention describes these cDNA molecules
CC   as useful for in hybridisation assays to detect expression of nucleic
CC   acids (or complementary nucleic acids) in a present in a given sample, as
CC   well as for screening assays by identifying molecules or compounds that
CC   specifically bind the cDNA as a ligand and modulate function or activity.
CC   Accordingly, these compositions exhibit cytostatic activity and can also
CC   be used for gene therapy purposes. This polynucleotide sequence is a cDNA
CC   that is differentially expressed in MYCN activated cells, given in an
CC   exemplification of the invention. NOTE: This sequence does not appear in
CC   the printed specification but has been obtained in electronic format from
CC   the US Patent Office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID=20030119009.
XX
SQ   Sequence 10432 BP; 3683 A; 2503 C; 2450 G; 1796 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               2.32e+03        Length:          10432
Score:                   9.00            Matches:         9
Percent Similarity:      100.0%          Conservative:    0
Best Local Similarity:   100.0%          Mismatches:      0
Query Match:             4.2%            Indels:          0
DB:                      10              Gaps:            0

US-09-989-890-238 (1-212) x ADJ56543 (1-10432)

Qy      29 AlaLeuSerArgProLeuSerPro 37
           |||||||||||||||||||||||||
Db    5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 70
ADE76855/c
ID   ADE76855 standard; cDNA; 10432 BP.
XX
AC   ADE76855;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human cDNA differentially expressed in a liver disorder #14.
XX
```

XX
KW human; ss; gene; liver disorder; hyperlipidaemia; hypertension;
KW type II diabetes; tumour; liver; inflammatory disorder;
KW immune response disorder; high-throughput screening;
KW differential gene expression; gene therapy.
XX
OS Homo sapiens.
XX
PN US2003108871-A1.
XX
PD 12-JUN-2003.
XX
PF 30-JUL-2001; 2001US-00919039.
XX
PR 28-JUL-2000; 2000US-0222113P.
XX
PA (KASE/) KASER M R.
XX
PI Kaser MR;
XX
DR WPI; 2004-031227/03.
DR P-PSDB; ADB76856.
XX
PT Composition comprising several cDNAs that are differentially expressed in
PT treated human C3A liver cell cultures, useful for treating liver
PT disorders.
XX
PS Claim 1; SEQ ID NO 20; 41pp; English.
XX
CC The invention relates to a composition comprising several cDNAs that are
CC differentially expressed in a liver disorder. The composition is useful
CC for treating liver disorder such as hyperlipidaemia, hypertension, type
CC II diabetes, tumours of the liver and disorders of the inflammatory and
CC immune response. The composition is useful for a high-throughput method
CC of screening several molecules or compounds to identify a ligand which
CC specifically binds a cDNA. A protein encoded by the cDNA is useful for a
CC high-throughput method for using a protein to screen several molecules or
CC compounds to identify at least one ligand which specifically binds the
CC protein which involves combining the protein encoded by the cDNA with
CC several of molecules or compounds under conditions to allow specific
CC binding, and detecting specific binding between the protein and a
CC molecule or compound, therefore identifying a ligand which specifically
CC binds the protein. The composition is useful for detecting and
CC quantifying differential gene expression, can be used in gene therapy, to
CC formulate prognosis and to design a treatment regimen and to monitor the
CC efficacy of treatment. The present sequence represents a cDNA
CC differentially expressed in a liver disorder.
XX
SQ Sequence 10432 BP; 3683 A; 2503 C; 2450 G; 1796 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.32e+03   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x ADB76855 (1-10432)

Qy   29 AlaLeuSerArgProLeuSerPro 37
              ||||||||||||||||||||||||
Db   5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 71
ADL12292/c
ID ADL12292 standard; cDNA; 10432 BP.
XX
AC ADL12292;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human steroid-induced C3A liver cell cDNA #21.

XX
KW ss; gene; Hepatotropic; Gene therapy; Wilson disease; liver disorder;
KW steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.
XX
OS Homo sapiens.
XX
PN US6673549-B1.
XX
PD 06-JAN-2004.
XX
PF 12-OCT-2001; 2001US-00976594.
XX
PR 12-OCT-2000; 2000US-0240409P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Furness LM, Buchbinder JL;
XX
DR WPI; 2004-068610/07.
XX
PT Combination useful for preparing a composition for treating liver
PT disorders associated with steroid therapy, e.g., cirrhosis or hepatitis,
PT comprises cDNAs that are differentially expressed in response to steroid
PT treatment.
XX
PS Claim 1; SEQ ID NO 21; 141pp; English.
XX
CC The invention relates to a combination comprising cDNAs that are
CC differentially expressed in response to steroid treatment. Also included
CC are the following: a high throughput method for using a cDNA to detect
CC differential expression of nucleic acids in a sample; and a high
CC throughput method of screening molecules or compounds to identify a
CC ligand that specifically binds a cDNA. The sample is from a subject with
CC Wilson disease and comparison of a standard defines a stage of that
CC disease. The high throughput method of screening molecules or compounds
CC to identify a ligand that specifically binds a cDNA comprises: combining
CC the combination with molecules or compounds under conditions to allow
CC specific binding; and detecting specific binding between each cDNA and at
CC least one molecule or compound. The molecules or compounds are regulatory
CC proteins. The combination is useful for preparing a composition for
CC treating liver disorders associated with steroid therapy, e.g. cirrhosis
CC or hepatitis. The present sequence represents a human C3A liver cells. Note: The
CC differentially expressed in steroid-induced C3A liver cells. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 10432 BP; 3683 A; 2503 C; 2450 G; 1796 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.32e+03   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x ADL12292 (1-10432)

Qy   29 AlaLeuSerArgProLeuSerPro 37
Db   5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 72
ADV85846/c
ID ADV85846 standard; cDNA; 10432 BP.
XX
AC ADV85846;
XX
DT 10-MAR-2005 (first entry)
XX
DE Human cDNA #75 differentially expressed in prostate cancer.
XX

KW Tumor marker; diagnostic; prostate cancer; cytostatic; gene therapy;
KW gene; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   128..9898
FT   /*tag= a
FT   /product= "Human protein differentially expressed in
FT   prostate cancer"
XX
PN US2004253609-A1.
XX
PD 16-DEC-2004.
XX
PF 06-JAN-2004; 2004US-00752986.
XX
PR 28-JUL-2000; 2000US-0222469P.
PR 30-JUL-2001; 2001US-00919172.
XX
PA (INCY-) INCYTE CORP.
XX
PI Faris M, Turner CM;
XX
DR WPI; 2005-056868/06.
DR P-PSDB; ADV85847.
XX
PT A composition comprising cDNAs that are differentially expressed in
PT prostate cancer, useful for diagnosing, staging or treating prostate
PT cancer or for monitoring the treatment of subjects with prostate cancer.
XX
PS Claim 1; SEQ ID NO 97; 144pp; English.
XX
CC The present invention relates to a composition comprising cDNAs that are
CC differentially expressed in prostate cancer. The composition is useful
CC for diagnosing, staging or treating prostate cancer or for monitoring the
CC treatment of a subject with prostate cancer. The invention is also used
CC in gene therapy. The present sequence is the human cDNA differentially
CC expressed in prostate cancer.
XX
SQ Sequence 10432 BP; 3683 A; 2503 C; 2450 G; 1796 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.32e+03   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 14   Gaps: 0

US-09-989-890-238 (1-212) x ADV85846 (1-10432)

Qy   29 AlaLeuSerArgProLeuSerPro 37
            ||||||||||||||||||||||||
Db   5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 73
ABS62800/c
ID ABS62800 standard; cDNA; 10433 BP.
XX
AC ABS62800;
XX
DT 05-NOV-2002 (first entry)
XX
DE Prostate adenocarcinoma associated cDNA #75.
XX
KW Prostate cancer; differential expression; cancer diagnosis;
KW cancer treatment; cancer monitoring; prostate adenocarcinoma; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002119463-A1.
XX

PD 29-AUG-2002.
XX
PF 30-JUL-2001; 2001US-00919172.
XX
PR 28-JUL-2000; 2000US-0222469P.
XX
PA (FARI/) FARIS M.
PA (TURN/) TURNER C M.
XX
PI Faris M, Turner CM;
XX
DR WPI; 2002-608155/65.
DR P-PSDB; ABG77188.
XX
PT New composition, useful for treating and diagnosing prostate cancer,
PT comprises human cDNAs that are differentially expressed in prostate
PT cancer.
XX
PS Claim 1; Page 130-135; 147pp; English.
XX
CC The invention describes a composition of a number of human cDNAs that are
CC differentially expressed in prostate cancer. The composition of the
CC invention useful for a high-throughput method for detecting differential
CC expression of cDNAs in a nucleic acid containing sample comprising
CC amplifying the nucleic acids of the sample, hybridising the composition
CC with nucleic acids of the sample, detecting the hybridisation complexes
CC and comparing the complexes with those of a standard, where differences
CC indicates differential expression. The sample is from a subject with
CC prostate cancer and comparison with a standard defines an early, mid or
CC late stage of the disease. The composition and or protein encoded by the
CC cDNA's are useful for a high-throughput method to screen molecules or
CC compounds to identify a ligand which specifically binds a cDNA or protein
CC comprising combining the composition or protein and detecting specific
CC binding between cDNA/protein and molecule or compound. The molecules and
CC compounds are selected from DNA, RNA, peptide nucleic acid molecules,
CC mimetics, peptides, proteins, (ant)agonists, antibodies, immunoglobulins,
CC inhibitors, drug compounds, pharmaceutical agents, transcription factors,
CC repressors, and regulatory proteins. The composition is useful for
CC diagnosing, treating or monitoring the progression or treatment of
CC prostate cancer. The antibodies are also useful for the diagnosis of
CC disease. This sequence represents a prostate adenocarcinoma cDNA
XX
SQ Sequence 10433 BP; 3683 A; 2503 C; 2450 G; 1796 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:   2.32e+03   Length: 10433
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 9
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x ABS62800 (1-10433)

Qy   29 AlaLeuSerArgProLeuSerPro 37
            ||||||||||||||||||||||||
Db   5064 GCTCTTCTTTCACGCCCACTTTCCCCA 5038

RESULT 74
AAA64659/c
ID AAA64659 standard; cDNA; 11435 BP.
XX
AC AAA64659;
XX
DT 02-JAN-2001 (first entry)
XX
DE cDNA sequence encoding the antigen of monoclonal antibody Ki-67.
XX
KW Melanoma; thyroid tumour; MEL3; MEL7; Thy5; Thy6; Thy11; Thy14; Thy15;
KW rectal cancer; lung cancer; breast cancer; colon cancer; antibody Ki-67;
KW ss.
XX
OS Homo sapiens.
XX

XX
PN WO20050595-A2.
XX
PD 31-AUG-2000.
XX
PF 25-FEB-2000; 2000WO-US004929.
XX
PR 25-FEB-1999;   99US-00257417.
XX
PA (GOUT/) GOUT I.
PA (RODN/) RODNIN N.
PA (FILO/) FILONENKO V.
PA (MATS/) MATSUKA G.
PA (SCAN/) SCANLAN M.
PA (OLDL/) OLD L.
PA (BILY/) BILYNSKY B.
XX
PI Gout I, Rodnin N, Filonenko V, Matsuka G, Scanlan M, Old L;
PI Bilynsky B;
XX
DR WPI; 2000-572092/53.
XX
PT Novel isolated nucleic acid molecules for diagnosing and treating
PT melanoma, thyroid tumors, rectal, lung, breast and colon cancers.
XX
PS Claim 17; Page 81-85; 94pp; English.
XX
CC The specification describes polynucleotides which are associated with
CC melanoma and thyroid tumors. Specifically, these are MEL3, MEL7, Thy5,
CC Thy6, Thy11, Thy14, and Thy15. The polynucleotides are useful for
CC diagnosing and treating a patient with melanoma, thyroid tumour, rectal
CC cancer, lung cancer, breast cancer or colon cancer. The present sequence
CC represents a polynucleotide of the invention
XX
SQ Sequence 11435 BP; 3802 A; 2828 C; 2684 G; 2121 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   11435
Score: 2.52e+03   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x AAA64659 (1-11435)

Qy   29 AlaLeuSerArgProLeuSerPro 37
                |||||||||||||||||||||||||
Db   4052 GCTCTTCTTTCACGCCCACTTTCCCCA 4026

RESULT 75
ADY16694/c
ID ADY16694 standard; DNA; 11435 BP.
XX
AC ADY16694;
XX
DT 05-MAY-2005 (first entry)
XX
DE DNA encoding a PRO polypeptide, SEQ ID NO 2500.
XX
KW Antiinflammatory; Immune disorder; Dermatological; Immunosuppressive;
KW Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;
KW Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;
KW Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;
KW Antiallergic; ds; gene; diagnosis.
XX
OS Homo sapiens.
XX
PN WO200501696962-A2.
XX
PD 24-FEB-2005.
XX
PF 11-AUG-2004; 2004WO-US026249.

XX
PR 11-AUG-2003; 2003US-0493546P.
XX
PA (GETH) GENENTECH INC.
XX
PI Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;
XX
DR WPI; 2005-182330/19.
XX
PT New nucleic acid encoding PRO polypeptide, useful for diagnosing and
PT treating an immune related disorder, e.g. systemic lupus erythematosus,
PT rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.
XX
PS Claim 1; SEQ ID NO 2500; 158pp; English.
XX
CC The invention relates to an isolated nucleic acid encoding a PRO
CC polypeptide. The polypeptide, agonist or an antagonist, antibody,
CC composition, and method are useful for diagnosing and treating an immune
CC related disorder, e.g. systemic lupus erythematosus, rheumatoid
CC arthritis. The present sequence represents a DNA encoding a PRO
CC polypeptide.
XX
SQ Sequence 11435 BP; 3802 A; 2828 C; 2684 G; 2121 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   11435
Score: 2.52e+03   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 14   Gaps: 0

US-09-989-890-238 (1-212) x ADY16694 (1-11435)

Qy   29 AlaLeuSerArgProLeuSerPro 37
                |||||||||||||||||||||||||
Db   4052 GCTCTTCTTTCACGCCCACTTTCCCCA 4026

RESULT 76
ADQ22788/c
ID ADQ22788 standard; DNA; 11466 BP.
XX
AC ADQ22788;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5608.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 5608; 210pp; English.

XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 11466 BP; 3830 A; 2830 C; 2684 G; 2122 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.53e+03   Length: 11466
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x ADQ22788 (1-11466)

Qy   29 AlaLeuSerArgProLeuSerPro 37
          ||||||||||||||||||||||||
Db   4052 GCTCTTCTTTCACGGCCCACTTCCCCA 4026

RESULT 77
AAZ43872/c
ID AAZ43872 standard; DNA; 12493 BP.
XX
AC AAZ43872;
XX
DT 10-MAR-2000 (first entry)
XX
DE Human cell cycle protein Ki-67 DNA.
XX
KW Cell cycle protein; Ki-67; therapy; cell proliferation; allergy;
KW tumor treatment; autoimmune disease; scar formation; inflammation;
KW rheumatic disease; transplantation; ds.
XX
OS Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   197..9967
FT   /*tag= a
FT   /product= "Ki-67"
XX
PN DE19822954-A1.
XX
PD 25-NOV-1999.
XX
PF 22-MAY-1998; 98DE-01022954.
XX
PR 22-MAY-1998; 98DE-01022954.
XX
PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
PI Flad H, Gerdes J, Boehle A, Deinert I;
XX
DR WPI; 2000-039964/04.
DR P-PSDB; AAY50976.
XX
PT Ki-67 gene antisense oligonucleotide.
XX
PS Disclosure; Page 5-19; 36pp; German.
XX
CC This invention describes a novel oligoribo- or oligodeoxyribonucleotide,
CC characterized in that, it hybridizes to mRNA that encodes protein Ki-67

XX
CC at a physiologically acceptable salt concentration. The oligoribo- or
CC oligodeoxyribonucleotide which is complementary to Ki-67, a protein
CC active at all stages of the cell cycle except G_0, is useful for therapy
CC of illnesses with increased cell proliferation and particularly for
CC treatment of tumors, autoimmune diseases, scar formation, inflammation,
CC allergy, rheumatic diseases and defence against transplantation. This
CC sequence encodes the human cell cycle protein Ki-67 which is described in
CC the method of the invention
XX
SQ Sequence 12493 BP; 4143 A; 3048 C; 2929 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.74e+03   Length: 12493
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x AAZ43872 (1-12493)

Qy   29 AlaLeuSerArgProLeuSerPro 37
          ||||||||||||||||||||||||
Db   5132 GCTCTTCTTTCACGGCCCACTTCCCCA 5106

RESULT 78
ADB53916/c
ID ADB53916 standard; cDNA; 12494 BP.
XX
AC ADB53916;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human prostate cancer cDNA #263.
XX
KW Human; prostate cancer; ss; cDNA combination; differential expression;
KW gene.
XX
OS Homo sapiens.
XX
PN US2003190640-A1.
XX
PD 09-OCT-2003.
XX
PF 29-MAY-2002; 2002US-00252157.
PR 31-MAY-2001; 2001US-0295048P.
XX
PA (FARI/) FARIS M.
PA (PEAR/) PEARSON C I.
XX
PI Faris M, Pearson CI;
XX
DR WPI; 2003-831619/77.
XX
PT New combination comprising cDNAs that are differentially expressed in
PT prostate cancer, useful for diagnosing, treating or monitoring the
PT progression of treatment of prostate cancer.
XX
PS Claim 1; SEQ ID NO 263; 42pp; English.
XX
CC The invention relates to a combination comprising a number of cDNAs
CC expressed in prostate cancer. The invention also relates to a method for
CC detecting differential expression of one or more cDNAs in a sample
CC containing nucleic acids by hybridising a substrate with the nucleic
CC acids, thus forming one or more hybridisation complexes, detecting
CC hybridisation complex formation and comparing the complexes formed with
CC standard complexes, where differences between the standard and the sample
CC complex formation indicate differential expression of cDNAs in the
CC sample. The differential expression is diagnostic of prostate cancer. The
CC invention also relates to proteins and antibodies related to the cDNAs.
CC The combination is useful for diagnosing, treating or monitoring the
CC progression of treatment of prostate cancer. The antibodies are useful

```
CC    for detecting prostate cancer. This sequence represents a human prostate
CC    cancer cDNA of the invention.
XX
SQ    Sequence 12494 BP; 4146 A; 3045 C; 2929 G; 2374 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.74e+03      Length:         12494
Score:                  9.00          Matches:        9
Percent Similarity:     100.0%        Conservative:   0       Percent Similarity:     100.0%    Conservative:   0
Best Local Similarity:  100.0%        Mismatches:     0       Best Local Similarity:  100.0%    Mismatches:     0
Query Match:            4.2%          Indels:         0       Query Match:            4.2%      Indels:         0
DB:                                   Gaps:           10      DB:                              Gaps:           3

US-09-989-890-238 (1-212) x ADE53916 (1-12494)                US-09-989-890-238 (1-212) x AAA64658 (1-12515)

Qy        29 AlaLeuSerArgProLeuSerPro 37                      Qy        29 AlaLeuSerArgProLeuSerPro 37
             ||||||||||||||||||||||||                                     ||||||||||||||||||||||||
Db      5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106                 Db      5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 79                                                     RESULT 80
AAA64658/c                                                    ABL69122/c
ID    AAA64658 standard; cDNA; 12515 BP.                      ID    ABL69122 standard; DNA; 12515 BP.
XX                                                            XX
AC    AAA64658;                                               AC    ABL69122;
XX                                                            XX
DT    02-JAN-2001  (first entry)                              DT    15-MAY-2002  (first entry)
XX                                                            XX
DE    cDNA sequence encoding the antigen of monoclonal antibody Ki-67.   DE    Kidney cancer related gene sequence SEQ ID NO:7459.
XX                                                            XX
KW    Melanoma; thyroid tumour; MEL3; MEL7; Thy5; Thy6; Thy11; Thy14; Thy15;   KW    Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW    rectal cancer; lung cancer; breast cancer; colon cancer; antibody Ki-67;   KW    stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW    ss.                                                     KW    cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX                                                            KW    gene; ds.
OS    Homo sapiens.                                           XX
XX                                                            OS    Homo sapiens.
PN    WO200050595-A2.                                         XX
XX                                                            PN    WO200194629-A2.
PD    31-AUG-2000.                                            XX
XX                                                            PD    13-DEC-2001.
PF    25-FEB-2000; 2000WO-US004929.                           XX
XX                                                            PF    30-MAY-2001; 2001WO-US010838.
PR    25-FEB-1999;  99US-00257417.                            XX
XX                                                            PR    05-JUN-2000; 2000US-0209473P.
PA    (GOUT/)  GOUT I.                                        PR    05-JUN-2000; 2000US-0209531P.
PA    (RODN/)  RODNIN N.                                      PR    18-SEP-2000; 2000US-0233133P.
PA    (FILO/)  FILONENKO V.                                   PR    18-SEP-2000; 2000US-0233617P.
PA    (MATS/)  MATSUKA G.                                     PR    20-SEP-2000; 2000US-0234009P.
PA    (SCAN/)  SCANLAN M.                                     PR    20-SEP-2000; 2000US-0234034P.
PA    (OLDL/)  OLD L.                                         PR    20-SEP-2000; 2000US-0234052P.
PA    (BILY/)  BILYNSKY B.                                    PR    22-SEP-2000; 2000US-0234509P.
XX                                                            PR    22-SEP-2000; 2000US-0234567P.
PI    Gout I, Rodnin N, Filonenko V, Matsuka G, Scanlan M, Old L;   PR    25-SEP-2000; 2000US-0234923P.
PI    Bilynsky B;                                             PR    25-SEP-2000; 2000US-0234924P.
XX                                                            PR    25-SEP-2000; 2000US-0235077P.
DR    WPI; 2000-572092/53.                                    PR    25-SEP-2000; 2000US-0235082P.
XX                                                            PR    25-SEP-2000; 2000US-0235134P.
PT    Novel isolated nucleic acid molecules for diagnosing and treating   PR    25-SEP-2000; 2000US-0235280P.
PT    melanoma, thyroid tumors, rectal, lung, breast and colon cancers.   PR    26-SEP-2000; 2000US-0235637P.
XX                                                            PR    26-SEP-2000; 2000US-0235638P.
PS    Claim 17; Page 77-81; 94pp; English.                    PR    27-SEP-2000; 2000US-0235711P.
XX                                                            PR    27-SEP-2000; 2000US-0235720P.
CC    The specification describes polynucleotides which are associated with   PR    27-SEP-2000; 2000US-0235840P.
CC    melanoma and thyroid tumors. Specifically, these are MEL3, MEL7, Thy5,   PR    27-SEP-2000; 2000US-0235863P.
CC    Thy6, Thy11, Thy14, and Thy15. The polynucleotides are useful for   PR    28-SEP-2000; 2000US-0236028P.
CC    diagnosing and treating a patient with melanoma, thyroid tumour, rectal   PR    28-SEP-2000; 2000US-0236032P.
CC    cancer, lung cancer, breast cancer or colon cancer. The present sequence   PR    28-SEP-2000; 2000US-0236033P.
CC    represents a polynucleotide of the invention                PR    28-SEP-2000; 2000US-0236034P.
XX                                                            PR    28-SEP-2000; 2000US-0236109P.
SQ    Sequence 12515 BP; 4164 A; 3050 C; 2928 G; 2373 T; 0 U; 0 Other;   PR    28-SEP-2000; 2000US-0236111P.
                                                              PR    29-SEP-2000; 2000US-0236842P.
Alignment Scores:                                             PR    29-SEP-2000; 2000US-0236891P.
Pred. No.:              2.74e+03      Length:         12515   PR    02-OCT-2000; 2000US-0237172P.
Score:                  9.00          Matches:        9       PR    02-OCT-2000; 2000US-0237173P.
                                                              PR    02-OCT-2000; 2000US-0237278P.
                                                              PR    02-OCT-2000; 2000US-0237294P.
                                                              PR    02-OCT-2000; 2000US-0237295P.
                                                              PR    03-OCT-2000; 2000US-0237316P.
                                                              PR    03-OCT-2000; 2000US-0237425P.
                                                              PR    03-OCT-2000; 2000US-0237598P.
                                                              PR    03-OCT-2000; 2000US-0237604P.
                                                              PR    03-OCT-2000; 2000US-0237606P.
```

PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 7459; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.74e+03   Length: 12515
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x ABL69122 (1-12515)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   5132 GCTCTTCTTCACGCCCACTTCCCCA 5106

RESULT 81
ADD12702/c
ID   ADD12702 standard; cDNA; 12515 BP.
XX
AC   ADD12702;
XX
DT   01-JAN-2004 (first entry)
XX
DE   Human cDNA expressed during adipocyte differentiation #48.
XX
KW   ss; gene; human; adipocyte differentiation; obesity; type II diabetes;
KW   lipodystrophy; hyperinsulinaemia.
XX
OS   Homo sapiens.
XX
PN   US2003113720-A1.
XX
PD   19-JUN-2003.
XX
PF   30-JUL-2001; 2001US-00918624.
XX
PR   28-JUL-2000; 2000US-0222470P.

XX
PA   (SCHE/) SCHEBYE X M.
PA   (SORN/) SORNASSE T.
XX
PI   Schebye XM, Sornasse T;
XX
DR   WPI; 2003-810888/76.
XX
PT   Novel isolated cDNAs expressed in adipocyte differentiation useful for
PT   treating subject with disorder such as obesity, type II diabetes,
PT   lipodystrophy or hyperinsulinemia.
XX
PS   Claim 1; SEQ ID NO 48; 105pp; English.
XX
CC   The invention relates to an isolated cDNA expressed during adipocyte
CC   differentiation. The cDNA is useful for treating a subject with a
CC   disorder such as obesity, type II diabetes, lipodystrophy or
CC   hyperinsulinaemia. The nucleic acid is useful for a high throughput
CC   method of using a cDNA to screen several molecules or compounds to
CC   identify a ligand which specifically binds the cDNA which involves
CC   combining the nucleic acid with several molecules or compounds under
CC   conditions to allow specific binding, and detecting specific binding
CC   between each cDNA and at least one molecule or compound, thus identifying
CC   a ligand that specifically binds to each cDNA. The several molecules or
CC   compounds are chosen from DNA molecules, RNA molecules, peptide nucleic
CC   acid molecules, mimetics, peptides, transcription factors, repressors and
CC   regulatory proteins. The present sequence represents a human cDNA
CC   expressed during adipocyte differentiation.
XX
SQ   Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.74e+03   Length: 12515
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x ADD12702 (1-12515)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   5132 GCTCTTCTTCACGCCCACTTCCCCA 5106

RESULT 82
ADG89365/c
ID   ADG89365 standard; DNA; 12515 BP.
XX
AC   ADG89365;
XX
DT   11-MAR-2004 (first entry)
XX
DE   Cancer detection method related gene #28.
XX
KW   ds; cancer; gene expression;
KW   estrogen receptor-positive invasive breast cancer.
XX
OS   Homo sapiens.
XX
PN   WO2003078662-A1.
XX
PD   25-SEP-2003.
XX
PF   12-MAR-2003; 2003WO-US007713.
XX
PR   13-MAR-2002; 2002US-0364890P.
PR   18-SEP-2002; 2002US-0412049P.
XX
PA   (GENO-) GENOMIC HEALTH INC.
XX
PI   Baker JB, Cronin MT, Kiefer MC, Shak S, Walker MG;
XX

DR   WPI; 2003-767536/72.
XX
PT   Predicting clinical outcome for a patient diagnosed with cancer comprises
PT   determining the expression level of one or more genes, and compared to
PT   the amount found in a reference cancer tissue set.
XX
PS   Disclosure; SEQ ID NO 313; 198pp; English.
XX
CC   The invention relates to a method of predicting clinical outcome for a
CC   patient diagnosed with cancer by determining the expression level of one
CC   or more genes, or their expression products, selected from p53BP2,
CC   cathepsin B, cathepsin L, Ki67/MiB1, and thymidine kinase in a cancer
CC   tissue obtained from the patient, normalized against control gene(s), and
CC   compared to the amount found in a reference cancer tissue set. The
CC   specification also discloses an array comprising polynucleotides
CC   hybridizing to the following genes: FOXM1, PRAME, Bcl2, STKI5, CEGP1, Ki-
CC   67, GSTM1, CA9, PR, BBC3, NME1, SURV, GATA3, TFRC, YB-1, DPYD, GSTM3,
CC   RPS6KB1, Sro, Chk1, ID1, EstR1, p27, CCNB1, XIAP, Chk2, CDC25B, IGFIR,
CC   AK055699, PI3KC2A, TGFB3, BAGI1, CYP3A4, EpCAM, VEGFC, pS2, hENT1, WISP1,
CC   HNF3A, NFKBp65, BRCA2, EGFR, TKI, VDR, ContigS51037, pENTI, EPHXI, IFIA,
CC   CDHI, HIFlt, IGFBP3, CTSB, Her2 and DIABLO, immobilized on a solid
CC   surface. The methods are useful for predicting clinical outcome for a
CC   patient diagnosed with cancer, classifying cancer, and predicting the
CC   likelihood of long-term survival of a breast cancer patient, or a patient
CC   diagnosed with invasive breast cancer or with estrogen receptor (ER)-
CC   positive invasive breast cancer. This sequence corresponds to a gene
CC   sequence whose expression is detected by the method of the invention.
XX
SQ   Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.74e+03   Length: 12515
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 10   Gaps: 0

US-09-989-890-238 (1-212) x ADG89365 (1-12515)

Qy   29 AlaLeuSerArgProLeuSerPro 37
          ||||||||||||||||||||||||||
Db   5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 83
ADN95435/c
ID   ADN95435 standard; DNA; 12515 BP.
XX
AC   ADN95435;
XX
DT   01-JUL-2004 (first entry)
XX
DE   Human BEC/LEC-related gene sequence SeqID358.
XX
KW   growth; differentiation; blood endothelial cell; BEC;
KW   lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
KW   lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;
KW   vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
KW   inflammatory disease; cancer metastasis; lymphatic system; gene; ds;
KW   human.
XX
OS   Homo sapiens.
XX
PN   WO2003080640-A1.
XX
PD   02-OCT-2003.
XX
PF   07-MAR-2003; 2003WO-US006900.
XX
PR   07-MAR-2002; 2002US-0363019P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
PA   (LICN) LICENTIA LTD.

XX
PI   Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
XX
DR   WPI; 2003-876899/81.
DR   P-PSDB; ADN95434.
XX
PS   Example 1; SEQ ID NO 358; 176pp; English.
XX
CC   This invention relates to a method of differentially modulating the
CC   growth or differentiation of blood endothelial cells (BEC) or lymphatic
CC   endothelial cells (LEC) comprises contacting endothelial cells with a
CC   composition comprising an agent that differentially modulates blood or
CC   lymphatic endothelial cells. Treating hereditary lymphoedema comprises
CC   identifying a human subject with lymphoedema and with a mutation in at
CC   least one allele of a gene encoding a LEC protein, where the mutation
CC   correlates with lymphoedema in human subjects, and with the proviso that
CC   the LEC protein is not VEGFR-3; and administering to the subject a
CC   composition comprising a lymphatic growth agent selected from VEGF-C or
CC   VEGF-D polypeptides and polynucleotides. The invention may be useful for
CC   the development of compounds with an antiangiogenic, cytostatic,
CC   vasotropic or antiinflammatory activity or for gene therapy. The method
CC   is useful in modulating the growth or differentiation of blood
CC   endothelial cells or lymphatic endothelial cells, in treating hereditary
CC   lymphoedema, in screening for an endothelial cell disorder or
CC   predisposition to the disorder or in monitoring the efficacy or toxicity
CC   of a drug on endothelial cells. The agent is useful in manufacturing a
CC   medicament for the differential modulation of blood vessel endothelial
CC   cell or lymphatic vessel endothelial cell growth or differentiation. The
CC   lymphatic growth agent may also be used in manufacturing a medicament for
CC   the treatment of hereditary lymphoedema resulting from a mutation in a
CC   LEC gene or of other diseases involving the lymphatic vessels, such as
CC   various inflammatory diseases and cancer metastasis via the lymphatic
CC   system. The present sequence is that of a human LEC/BEC differentially
CC   expressed gene which is related to the method of the invention.Note:This
CC   sequence does not appear in the specification but was obtained by the
CC   indexer using the source data given in table 14 of the specification.
XX
SQ   Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.74e+03   Length: 12515
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 11   Gaps: 0

US-09-989-890-238 (1-212) x ADN95435 (1-12515)

Qy   29 AlaLeuSerArgProLeuSerPro 37
          ||||||||||||||||||||||||||
Db   5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 84
ADU06058/c
ID   ADU06058 standard; DNA; 12515 BP.
XX
AC   ADU06058;
XX
DT   27-JAN-2005 (first entry)
XX
DE   Novel bronchial cancer-associated human gene SeqID280.
XX
KW   bronchial cancer; cytostatic; tumour-associated protein;
KW   cancer detection; metastasis; tumour; gene; ds; human.
XX
OS   Homo sapiens.
XX
PN   DE10316701-A1.
XX
PD   04-NOV-2004.
XX
PF   09-APR-2003; 2003DE-01016701.

```
XX
PR   09-APR-2003; 2003DE-01016701.
XX
PA   (HINZ/)  HINZMANN B.
PA   (HERM/)  HERMANN K.
PA   (CAST/)  HEIDEN CASTANOS-VELEZ E.
XX
PI   Mennerich D, Bruemmendorf T, Heiden E, Hermann K, Kinnemann H;
PI   Li X, Roepcke S, Staub E, Hinzmann B, Rosenthal A, Pilarsky C;
XX
DR   WPI; 2004-786403/78.
DR   P-PSDB; ADU06545.
XX
PT   New nucleic acid, and derived proteins, useful for diagnosis of bronchial
PT   cancer and in screening for therapeutic and diagnostic agents.
XX
PS   Claim 1; SEQ ID NO 280; 1381pp; German.
XX
CC   This invention relates to a novel isolated nucleic acid associated with
CC   bronchial cancer comprising 489 defined sequences given in the
CC   specification. The invention may be useful for the production of
CC   compounds with a cytostatic activity through the inhibition of expression
CC   or activity of tumour-associated proteins. The novel DNA sequences and
CC   the proteins/peptides encoded by them are used for developing bronchial
CC   cancer or determining the risk of developing it and to screen for
CC   specific binding partners of the DNA or protein sequences, where the
CC   binding partners are potentially useful as agents for treating or
CC   diagnosing bronchial cancer. The DNA or protein sequences can also be
CC   used for prognosis, detection of metastases and for secondary treatment
CC   (of tumours that have been stabilised or are no longer detectable).
CC   Detecting abnormal expression of the DNA sequences provides early
CC   diagnosis of bronchial cancers. The present sequence is that of a novel
CC   bronchial cancer-associated human gene sequence of the invention.
XX
SQ   Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                              Length:        12515
Score:                2.74e+03          Matches:           9
Percent Similarity:   9.00              Conservative:      0
Best Local Similarity: 100.0%           Mismatches:        0
Query Match:          100.0%            Indels:            0
                      4.2%              Gaps:              0
                      13

US-09-989-890-238 (1-12515) x ADU06058 (1-12515)

Qy     29 AlaLeuSerArgProLeuSerPro 37
          ||||||||||||||||||||||||
Db   5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 85
ADX08067/c
ID   ADX08067 standard; DNA; 12515 BP.
XX
AC   ADX08067;
XX
DT   21-APR-2005 (first entry)
XX
DE   Cyclin-dependent kinase modulation biomarker DNA SEQ ID NO 2632.
XX
KW   cytostatic; cyclin-dependent kinase; cdk; biomarker; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2005012875-A2.
XX
PD   10-FEB-2005.
XX
PF   29-JUL-2004; 2004WO-US024424.
XX
PR   29-JUL-2003; 2003US-0490890P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
```

```
XX
PI   Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;
XX
DR   WPI; 2005-163068/17.
DR   P-PSDB; ADX08068.
XX
PT   Biomarkers useful for predicting or determining the response of a mammal
PT   to a cancer treatment comprising administration of a modulator of cyclin-
PT   dependent kinase activity.
XX
PS   Claim 5; SEQ ID NO 2632; 141pp; English.
XX
CC   This invention describes a novel method of predicting or determining
CC   whether a mammal will respond or is responding to an anti-cancer agent
CC   that modulates cyclin-dependent kinase (cdk) activity. The method
CC   comprises measuring the level of one or more biomarkers selected from
CC   2774 biomarkers given in the specification (nucleotide sequence SEQ ID
CC   NO:1246 (Genbank EST W28729) is especially preferred). The method of the
CC   invention is utilized in a kit for determining or predicting whether
CC   patient would be susceptible or resistant to treatment by an agent
CC   modulating cdk activity. The invention also describes a method for
CC   utilizing individualized genetic profiles for treating diseases and
CC   disorders based on patient's response and molecular level, specialized
CC   microarrays comprising the biomarkers described, antibodies directed
CC   against the biomarkers and a cell culture model to identify biomarkers.
CC   The cdk modulator is preferably N-5-[5-(1,1-Dimethylethyl)-2-
CC   oxazolyl]methyl]thio]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-
CC   tartaric acid salt. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct sequences. This
CC   sequence encodes a biomarker used in the method of the invention.
XX
SQ   Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                              Length:        12515
Score:                2.74e+03          Matches:           9
Percent Similarity:   9.00              Conservative:      0
Best Local Similarity: 100.0%           Mismatches:        0
Query Match:          100.0%            Indels:            0
                      4.2%              Gaps:              0
                      14

US-09-989-890-238 (1-212) x ADX08067 (1-12515)

Qy     29 AlaLeuSerArgProLeuSerPro 37
          ||||||||||||||||||||||||
Db   5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 86
ADY16692/c
ID   ADY16692 standard; DNA; 12515 BP.
XX
AC   ADY16692;
XX
DT   05-MAY-2005 (first entry)
XX
DE   DNA encoding a PRO polypeptide, SEQ ID NO 2498.
XX
KW   Antiinflammatory; Immune disorder; Dermatological; Immunosuppressive;
KW   Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;
KW   Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;
KW   Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;
KW   Antiallergic; ds; gene; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO2005016962-A2.
XX
PD   24-FEB-2005.
XX
PF   11-AUG-2004; 2004WO-US026249.
XX
PR   11-AUG-2003; 2003US-0493546P.
```

```
XX
PA  (GETH ) GENENTECH INC.
XX
PI  Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;
XX
DR  WPI; 2005-182330/19.
XX
PT  New nucleic acid encoding PRO polypeptide, useful for diagnosing and
PT  treating an immune related disorder, e.g. systemic lupus erythematosus,
PT  rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.
XX
PS  Claim 1; SEQ ID NO 2498; 158pp; English.
XX
CC  The invention relates to an isolated nucleic acid encoding a PRO
CC  polypeptide. The polypeptide, agonist or an antagonist, antibody,
CC  composition, and method are useful for diagnosing and treating an immune
CC  related disorder, e.g. systemic lupus erythematosus, rheumatoid
CC  arthritis. The present sequence represents a DNA encoding a PRO
CC  polypeptide.
XX
SQ  Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.74e+03    Length:          12515
Score:                  9.00        Matches:         9
Percent Similarity:     100.0%      Conservative:    0
Best Local Similarity:  100.0%      Mismatches:      0
Query Match:            4.2%        Indels:          0
DB:                     14          Gaps:            0

US-09-989-890-238 (1-212) x ADY16692 (1-12515)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db  5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 87
ADY16690/c
ID  ADY16690 standard; DNA; 12515 BP.
XX
AC  ADY16690;
XX
DT  05-MAY-2005  (first entry)
XX
DE  DNA encoding a PRO polypeptide, SEQ ID NO 2496.
XX
KW  Antiinflammatory; Immune disorder; Dermatological; Immunosuppressive;
KW  Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;
KW  Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;
KW  Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;
KW  Antiallergic; ds; gene; diagnosis.
XX
OS  Homo sapiens.
XX
PN  WO200516962-A2.
XX
PD  24-FEB-2005.
XX
PF  11-AUG-2004; 2004WO-US026249.
PR  11-AUG-2003; 2003US-0493546P.
XX
PA  (GETH ) GENENTECH INC.
XX
PI  Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;
XX
DR  WPI; 2005-182330/19.
XX
PT  New nucleic acid encoding PRO polypeptide, useful for diagnosing and
PT  treating an immune related disorder, e.g. systemic lupus erythematosus,
PT  rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.
XX
PS  Claim 1; SEQ ID NO 2496; 158pp; English.
```

```
XX
CC  The invention relates to an isolated nucleic acid encoding a PRO
CC  polypeptide. The polypeptide, agonist or an antagonist, antibody,
CC  composition, and method are useful for diagnosing and treating an immune
CC  related disorder, e.g. systemic lupus erythematosus, rheumatoid
CC  arthritis. The present sequence represents a DNA encoding a PRO
CC  polypeptide.
XX
SQ  Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.74e+03    Length:          12515
Score:                  9.00        Matches:         9
Percent Similarity:     100.0%      Conservative:    0
Best Local Similarity:  100.0%      Mismatches:      0
Query Match:            4.2%        Indels:          0
DB:                     14          Gaps:            0

US-09-989-890-238 (1-212) x ADY16690 (1-12515)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db  5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 88
ADY61174/c
ID  ADY61174 standard; DNA; 12515 BP.
XX
AC  ADY61174;
XX
DT  19-MAY-2005  (first entry)
XX
DE  Breast cancer associated human polynucleotide sequence, SEQ ID No:58.
XX
KW  gene expression; breast tumor; ERBB2; HER2; Neu; oncogene; chromosome 17;
KW  DNA library; DNA microarray; diagnosis; cytostatic; ds.
XX
OS  Homo sapiens.
XX
PN  WO2005021788-A1.
XX
PD  10-MAR-2005.
XX
PF  27-AUG-2004; 2004WO-IB002968.
PR  28-AUG-2003; 2003US-0498497P.
PR  27-AUG-2004; 2004US-00928465.
XX
PA  (IPSO-) IPSOGEN.
PA  (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA  (PAOL-) INST PAOLI CALMETTES.
XX
PI  Birnbaum D, Bertucci F, Jacquemier J, Debono S, Borie N;
PI  Ginestier C;
XX
DR  WPI; 2005-223195/23.
XX
PT  Analyzing differential gene expression, useful for detecting, diagnosing,
PT  staging, monitoring, predicting, preventing or treating breast cancer,
PT  comprises detecting an ERBB2 gene expression signature.
XX
PS  Claim 5; SEQ ID NO 58; 176pp; English.
XX
CC  The invention relates to a method of analyzing differential gene
CC  expression associated with breast tumor. The method comprises detecting
CC  the over expression or under expression of sequences neighboring the
CC  ERBB2 (also known as HER2 or Neu) oncogene on chromosome 17q12. The
CC  polynucleotide expression profiling comprises at least one, preferably at
CC  least two, three or all polynucleotide sequences (or its subsequence or
CC  complement) selected from predefined polynucleotide sequence sets
CC  consisting of: Set 1: SEQ ID NOS: 73-77 (ERBB2), Set 2: SEQ ID NOS: 28-30
CC  (GRB7), Set 3: SEQ ID NOS: 83-85 (NR1D1), Set 4: SEQ ID NOS: 78-80
CC  (GATA4), Set 5: SEQ ID NOS: 41-43 (CDH15), Set 6: SEQ ID NOS: 16, 17
```

```
CC    (LTA), Set 7: SEQ ID NOS: 86, 87, 116 (MAP2K6), Set 8: SEQ ID NOS: 54,
CC    55, 113 (PECAM1), Set 9: SEQ ID NOS: 44, 45 (PPARBP), Set 13: SEQ ID NO:
CC    10 (LOC148696), Set 18: SEQ ID NO: 24, 25 (STAT3), Set 20: SEQ ID NOS: 36
CC    -38 (CDKL5), Set 21: SEQ ID NO: 46-48 (CSTA), Set 22: SEQ ID NOS: 52,
CC    53, 115 (ITGB3), Set 23: SEQ ID NOS: 56-58 (MKI67), Set 24: SEQ ID NOS:
CC    59-61 (PBEF), Set 27: SEQ ID NOS: 88-90 (ITGA2), Set 28: SEQ ID NO: 11
CC    (ESTAA878915), SET 29: SEQ ID NOS: 1-3 (JDP1), SET 35: SEQ ID NOS: 67-69
CC    (FLJ10193), SET 36: SEQ ID NOS: 70-72 (ESR1), SET 43: SEQ ID NOS: 104-106
CC    (DAXX), SET 47: SEQ ID NO: 114, or SET 48: SEQ ID NOS: 117, 118
CC    (C17ORF37). Also described are: (i) a method for monitoring the treatment
CC    of a patient with a breast cancer, (ii) a polynucleotide library, useful
CC    for the molecular characterization of a breast cancer comprising or
CC    corresponding to a pool of polynucleotide sequences, over or under
CC    expressed, in breast tissue, the pool corresponding to the polynucleotide
CC    sequences defined above, and (iii) a method for treating a patient with
CC    breast cancer. The method further comprises detecting the over expression
CC    of at least one, preferably at least two or three, polynucleotide
CC    sequence (or its subsequence or complement) from predefined
CC    polynucleotide sequence sets consisting of: Set 10: SEQ ID NOS: 33-35
CC    (PPP1R1B), Set 11: SEQ ID NOS: 39, 40 (RPL19), Set 12: SEQ ID NO. 4, 5, 6
CC    (PSMB3), Set 14: SEQ ID NOS: 12, 13 (NOL3/loc283849), Set 15: SEQ ID NOS:
CC    14, 15 (ITGA2B), Set 16: SEQ ID NOS: 18, 19 (NFKBIE), Set 17: SEQ ID NOS:
CC    22, 23 (PADI2), Set 19: SEQ ID NOS: 26, 27 (OAS2), Set 25: SEQ ID NOS: 62
CC    -64 (FADS2), or Set 26: SEQ ID NOS: 81, 82 (LOX). The method further
CC    comprises detecting the under expression of at least one, preferably at
CC    least two or three, polynucleotide sequence (or its subsequence or
CC    complement) from predefined polynucleotide sequence sets consisting of:
CC    SET 30: SEQ ID NOS: 7-9 (NA), SET 31: SEQ ID NOS: 20, 21 (CELSR2), SET
CC    32: SEQ ID NOS: 31, 32 (ESTN33243/NA), SET 33: SEQ ID NOS: 49-51
CC    (SCUBE2), or SET 34: SEQ ID NOS: 65, 66 (BSTH29301/NA). The differential
CC    gene expression corresponds to an alteration of ERBB2 or ER gene
CC    expression in breast tumor. The detection of over or under expression of
CC    polynucleotide sequences is carried out by FISH or IHC. The detection of
CC    over or under expression of polynucleotide sequences is performed on
CC    nucleic acids from a breast tissue sample or from a tumor cell line, such
CC    as a breast cell line. The detection of over or under expression of
CC    polynucleotide sequences is performed on DNA microarrays. The detection
CC    of over or under expression of polynucleotide sequences is carried out at
CC    the protein level, where the detection is performed on proteins expressed
CC    from nucleic acid in a breast tissue sample or cell line. The method
CC    comprises (a) obtaining nucleic acids from a breast tissue sample from a
CC    patient, and (b) reacting the nucleic acids sample obtained in step (a)
CC    with a polynucleotide library, and (c) detecting the reaction product of
CC    step (b). The polynucleotide library is immobilized on a solid support,
CC    where the support is nylon membrane, nitrocellulose membrane, glass
CC    slide, glass beads, membranes on glass support or silicon chip. The
CC    method of the invention is useful for the identification of an ERBB2 gene
CC    expression signature in breast cancer. The methods and sequences are
CC    useful for detecting, diagnosing, staging, monitoring, predicting,
CC    preventing or treating conditions associated with breast cancer, and are
CC    used to follow up the stage or aggressiveness of a breast cancer. This
CC    sequence represents a human polynucleotide sequence located on chromosome
CC    17q12 that can be used in the method of the invention.
XX
SQ    Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                  2.74e+03         Length:             12515
Score:                      9.00             Matches:            9
Percent Similarity:         100.0%           Conservative:       0
Best Local Similarity:      100.0%           Mismatches:         0
Query Match:                4.2%             Indels:             0
DB:                         14               Gaps:               0

US-09-989-890-238 (1-212) x ADY61174 (1-12515)

Qy         29 AlaLeuSerArgProLeuSerPro 37
              ||||||||||||||||||||||||||
Db       5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 89
AD209648/c
ID   AD209648 standard; DNA; 12515 BP.
```

```
XX
AC    AD209648;
XX
DT    30-JUN-2005   (first entry)
XX
DE    Human breast cancer marker MKI67 DNA.
XX
KW    neoplasm; tumor marker; cytostatic; breast tumor; gene therapy; ds; gene;
KW    MKI67.
XX
OS    Homo sapiens.
XX
PN    EP1522594-A2.
XX
PD    13-APR-2005.
XX
PF    30-JUN-2004; 2004EP-00015374.
XX
PR    06-OCT-2003; 2003EP-00022587.
XX
PA    (FARB ) BAYER HEALTHCARE AG.
XX
PI    Munnes M, Bojar H;
XX
DR    WPI; 2005-265481/28.
DR    P-PSDB; AD209813.
DR    REFSEQ; NM_002417.
XX
PT    Characterizing the state of a neoplastic disease in a subject by
PT    comparing the pattern of determined expression levels of marker genes in
PT    a biological sample with reference patterns of expression levels.
XX
PS    Claim 1; SEQ ID NO 92; 542pp; English.
XX
CC    The invention relates to a novel method for characterizing the state of a
CC    neoplastic disease in a subject comprising determining the pattern of
CC    expression levels of at least 6, 8,10,15, 20, 30, 47 or 67 marker genes
CC    in a biological sample, comparing the pattern of expression levels with
CC    reference patterns of expression levels and characterizing the state of
CC    the neoplastic disease in the subject from the outcome of the comparison.
CC    The method of the invention demonstrates cytostatic applications and may
CC    be useful in characterizing the state of a neoplastic disease in a
CC    subject, preferably breast cancer, in gene therapy and for screening for
CC    compounds having a curative effect on a neoplastic disease. The current
CC    sequence is that of the human breast cancer marker MKI67 DNA of the
CC    invention which is differentially expressed in neoplastic tissue.
XX
SQ    Sequence 12515 BP; 4166 A; 3048 C; 2928 G; 2373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                  2.74e+03         Length:             12515
Score:                      9.00             Matches:            9
Percent Similarity:         100.0%           Conservative:       0
Best Local Similarity:      100.0%           Mismatches:         0
Query Match:                4.2%             Indels:             0
DB:                         14               Gaps:               0

US-09-989-890-238 (1-212) x ADZ09648 (1-12515)

Qy         29 AlaLeuSerArgProLeuSerPro 37
              ||||||||||||||||||||||||||
Db       5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 90
ACA03928/c
ID   ACA03928 standard; cDNA; 12632 BP.
XX
AC    ACA03928;
XX
DT    27-MAY-2003   (first entry)
XX
DE    cDNA downregulated in senescent cells Incyte ID NO: 412661.2.
XX
```

KW   Human; senescence; ss; gene; cancer; proliferative disorder; leukaemia;
KW   adenocarcinoma; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   adrenal gland cancer; bladder cancer; bone marrow cancer; heart cancer;
KW   brain cancer; breast cancer; cervical cancer; colon cancer; heart cancer;
KW   oesophageal cancer; gall bladder cancer; ganglial cancer; kidney cancer;
KW   liver cancer; lung cancer; muscle cancer; ovarian cancer; penile cancer;
KW   pancreatic cancer; parathyroid gland cancer; prostate cancer;
KW   salivary gland cancer; skin cancer; small intestine cancer;
KW   spleen cancer; stomach cancer; testicular cancer; thymic cancer;
KW   thyroid cancer; uterine cancer.
XX
OS   Homo sapiens.
XX
PN   US2002192678-A1.
XX
PD   19-DEC-2002.
XX
PF   07-FEB-2002; 2002US-00071766.
XX
PR   09-FEB-2001; 2001US-0268380P.
XX
PA   (CHEN/) CHEN H.
XX
PI   Chen H;
XX
DR   WPI; 2003-328858/31.
XX
PT   New combination comprising cDNAs or their complements, useful for
PT   detecting changes in expression of genes encoding proteins associated
PT   with senescence, and in diagnosing, staging or treating proliferative
PT   diseases, e.g. cancer.
XX
PS   Example 13; Page 71-76; 195pp; English.
XX
CC   The invention relates to a combination comprising a plurality of cDNAs,
CC   or their complements that are differentially expressed in cancer and
CC   other proliferative disorders. The combination is useful in detecting
CC   changes in expression of genes encoding proteins that are associated with
CC   senescence and in diagnosing, staging, treating, or monitoring the
CC   progression or treatment of subjects with proliferative diseases such as
CC   cancer e.g. adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma,
CC   sarcoma, teratocarcinoma; cancer of the adrenal gland, bladder, bone,
CC   bone marrow, brain, breast, cervix, colon, oesophagus, gall bladder,
CC   ganglia, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid
CC   gland, penis, prostate, salivary glands, skin, small intestine, spleen,
CC   stomach, testis, thymus, thyroid and uterus. The present sequence
CC   represents cDNA of genes that are downregulated in senescent cells
XX
SQ   Sequence 12632 BP; 4157 A; 3089 C; 2999 G; 2387 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.77e+03   Length: 12632
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x ACA03928 (1-12632)

Qy   29 AlaLeuSerArgProLeuSerPro 37
             ||||||||||||||||||||||||
Db   5270 GCTCTTCTTTCACGCCCACTTTCCCCA 5244

RESULT 91
ACN39127/c
ID   ACN39127 standard; cDNA; 12674 BP.
XX
AC   ACN39127;
XX
DT   18-NOV-2004 (first entry)
XX
DE   Tumour-associated antigenic target (TAT) cDNA DNA325386, SEQ ID NO:3074.

KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2004030615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Wu TD,   Zhang Z,   Zhou Y;
XX
DR   WPI; 2004-347921/32.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 1; SEQ ID NO 3074; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT nucleic acid of the invention
XX
SQ   Sequence 12674 BP; 4164 A; 3103 C; 3017 G; 2390 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   2.77e+03   Length: 12674
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x ACN39127 (1-12674)

Qy   29 AlaLeuSerArgProLeuSerPro 37
             ||||||||||||||||||||||||
Db   5312 GCTCTTCTTTCACGCCCACTTTCCCCA 5286

RESULT 92
ADF81724/c
ID   ADF81724 standard; DNA; 12747 BP.
XX

```
AC   ADF81724;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Leukaemia-related DNA sequence #2280.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
XX
OS   Unidentified.
XX
PN   WO200303943-A2.
XX
PD   15-MAY-2003.
XX
PF   04-NOV-2002; 2002WO-EP012303.
XX
PR   05-NOV-2001; 2001EP-00126244.
PR   30-APR-2002; 2002EP-00009758.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA   (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA   (HAFE/) HAFERLACH T.
PA   (SCHO/) SCHOCH C.
PA   (KERN/) KERN W.
XX
PI   Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI   Eils R, Brors B, Mergenthaler S;
XX
DR   WPI; 2003-505037/47.
XX
PT   Determining the subtype of leukemia cells and whether a patient sample
PT   contains leukemia cells or other cells, useful for treating leukemia,
PT   comprises determining the expression profile of a group of markers in a
PT   patient sample.
XX
PS   Disclosure; SEQ ID NO 2280; 2938pp; English.
XX
CC   The present invention relates to a method (M1) for determining the
CC   subtype of leukaemia cells and whether a patient sample contains
CC   leukaemia cells. The method comprises determining the expression profile
CC   of a group of markers in a patient sample. The method is useful for
CC   determining the presence of leukaemia cells, its types or subtypes, and
CC   for the preparation of a medicament for treating leukaemia.
XX
SQ   Sequence 12747 BP; 4164 A; 3071 C; 2985 G; 2365 T; 0 U; 162 Other;

Alignment Scores:
Pred. No.:                          Length:          12747
Score:                 2.79e+03     Matches:         9
                       9.00         Conservative:    0
Percent Similarity:    100.0%       Mismatches:      0
Best Local Similarity: 100.0%       Indels:          0
Query Match:           4.2%         Gaps:            0
DB:                    10

US-09-989-890-238 (1-212) x ADF81724 (1-12747)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db  5326 GCTCTTCTTTCACGCCCACTTTCCCCA 5300

RESULT 93
ADF81723/c
ID   ADF81723 standard; DNA; 12747 BP.
XX
AC   ADF81723;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Leukaemia-related DNA sequence #2279.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
XX
OS   Unidentified.
```

```
AC   ADF81724;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Leukaemia-related DNA sequence #2280.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
XX
OS   Unidentified.
XX
PN   WO200303943-A2.
XX
PD   15-MAY-2003.
XX
PF   04-NOV-2002; 2002WO-EP012303.
XX
PR   05-NOV-2001; 2001EP-00126244.
PR   30-APR-2002; 2002EP-00009758.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA   (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA   (HAFE/) HAFERLACH T.
PA   (SCHO/) SCHOCH C.
PA   (KERN/) KERN W.
XX
PI   Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI   Eils R, Brors B, Mergenthaler S;
XX
DR   WPI; 2003-505037/47.
XX
PT   Determining the subtype of leukemia cells and whether a patient sample
PT   contains leukemia cells or other cells, useful for treating leukemia,
PT   comprises determining the expression profile of a group of markers in a
PT   patient sample.
XX
PS   Disclosure; SEQ ID NO 2279; 2938pp; English.
XX
CC   The present invention relates to a method (M1) for determining the
CC   subtype of leukaemia cells and whether a patient sample contains
CC   leukaemia cells. The method comprises determining the expression profile
CC   of a group of markers in a patient sample. The method is useful for
CC   determining the presence of leukaemia cells, its types or subtypes, and
CC   for the preparation of a medicament for treating leukaemia.
XX
SQ   Sequence 12747 BP; 4164 A; 3071 C; 2985 G; 2365 T; 0 U; 162 Other;

Alignment Scores:
Pred. No.:                          Length:          12747
Score:                 2.79e+03     Matches:         9
                       9.00         Conservative:    0
Percent Similarity:    100.0%       Mismatches:      0
Best Local Similarity: 100.0%       Indels:          0
Query Match:           4.2%         Gaps:            0
DB:                    10

US-09-989-890-238 (1-212) x ADF81723 (1-12747)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db  5326 GCTCTTCTTTCACGCCCACTTTCCCCA 5300

RESULT 94
ADG88601
ID   ADG88601 standard; DNA; 48396 BP.
XX
AC   ADG88601;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Human PROX-1 DNA #2.
XX
KW   ds; gene; human; prox-1; developmental disorder; Usher Syndrome Type II;
KW   ocular disorder; retinal degradation; retinitis pigmentosa.
XX
OS   Homo sapiens.
XX
PN   US2003224516-A1.
XX
PD   04-DEC-2003.
XX
PF   03-JUN-2002; 2002US-00162846.
XX
PR   03-JUN-2002; 2002US-00162846.
```

```
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Dobie KW;
XX
DR   WPI; 2004-022080/02.
XX
PT   New compounds, particularly antisense oligonucleotides targeted to a
PT   nucleic acid encoding prox-1 useful for treating a disease or condition
PT   associated with prox-1, e.g. cancer, Alzheimer's disease or
PT   neurodegenerative disease.
XX
PS   Example 15; SEQ ID NO 11; 99pp; English.
XX
CC   The invention relates to a compound targeted to a nucleic acid molecule
CC   encoding prox-1 and inhibits the expression of prox-1. The compound,
CC   composition and methods are useful for treating a disease or condition
CC   associated with prox-1, such as a developmental disorder e.g. Usher
CC   Syndrome Type II, or an ocular disorder, e.g. retinal degradation or
CC   retinitis pigmentosa. They are also useful in research and diagnostics
CC   for modulating the expression of prox-1. The present sequence represents
CC   a human PROX-1 DNA.
XX
SQ   Sequence 48396 BP; 13584 A; 9992 C; 9850 G; 14970 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:             9.54e+03       Length:          48396
Score:                 9.00           Matches:         9
Percent Similarity:    100.0%         Conservative:    0
Best Local Similarity: 100.0%         Mismatches:      0
Query Match:           4.2%           Indels:          0
DB:                    12             Gaps:            0

US-09-989-890-238 (1-212) x ADG88601 (1-48396)

Qy     24 LeuProSerLeuAlaLeuLeuSer 32
          ||||||||||||||||||||||||||
Db     75 CTCCCAAGTTCTCTTGCCTTGCTATCC 101

RESULT 95
ADX70387
ID   ADX70387 standard; DNA; 49275 BP.
XX
AC   ADX70387;
XX
DT   05-MAY-2005  (first entry)
XX
DE   Human Prox-1 genomic DNA.
XX
KW   colorectal tumor; neoplasm; Prox-1; Cytostatic; ds; gene.
XX
OS   Homo sapiens.
XX
PN   WO2005014854-A1.
XX
PD   17-FEB-2005.
XX
PF   06-AUG-2004; 2004WO-EP008819.
XX
PR   08-AUG-2003; 2003US-0494221P.
XX
PA   (LICN ) LICENTIA LTD.
XX
PI   Alitalo K, Petrova T, Nykanen A;
XX
DR   WPI; 2005-152553/16.
DR   P-PSDB; ADX70389.
XX
PT   Screening, diagnosis and therapy of colon tissue for a pathological
PT   condition, e.g. colorectal cancer, where elevation of transcription
PT   factor Prox-1 expression correlates with a pathological phenotype.
XX
PS   Claim 71; SEQ ID NO 1; 270pp; English.
XX
CC   The invention relates to a method of screening colon tissue for a
CC   pathological condition which comprises measuring Prox-1 expression in a
CC   biological sample that comprises colon tissue from a mammalian subject,
CC   where elevated Prox-1 expression in the colon tissue correlates with a
CC   pathological phenotype. The molecule that suppressed expression or
CC   activity of Prox-1 is useful in the manufacture of a medicament for the
CC   treatment of colorectal cancer. The inhibitor of Prox-1 function in
CC   mammalian cells is useful for the manufacture for inhibiting Prox-1
CC   function. The siRNA molecule is useful in the manufacture of a medicament
CC   for the treatment of colorectal cancer. The present sequence represents
CC   the human Prox-1 genomic DNA.
XX
SQ   Sequence 49275 BP; 13857 A; 10175 C; 10034 G; 15209 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:             9.7e+03        Length:          49275
Score:                 9.00           Matches:         9
Percent Similarity:    100.0%         Conservative:    0
Best Local Similarity: 100.0%         Mismatches:      0
Query Match:           4.2%           Indels:          0
DB:                    14             Gaps:            0

US-09-989-890-238 (1-212) x ADX70387 (1-49275)

Qy     24 LeuProSerLeuAlaLeuLeuSer 32
          ||||||||||||||||||||||||||
Db    276 CTCCCAAGTTCTCTTGCCTTGCTATCC 302

RESULT 96
ABS56296
ID   ABS56296 standard; cDNA; 73544 BP.
XX
AC   ABS56296;
XX
DT   17-JAN-2003  (first entry)
XX
DE   Human transporter protein gene.
XX
KW   Human; gene; ds; transporter protein; allelic variant; SNP;
KW   cell proliferation; cell differentiation; cell signalling; antibody;
KW   gene chip; transgenic; therapeutic; diagnostic;
KW   single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            3000..70546
FT                  /*tag= a
FT                  /product= "transporter protein"
FT   exon           3000..3401
FT                  /*tag= b
FT                  /number= 1
FT   intron         3402..4087
FT                  /*tag= c
FT                  /number= 1
FT                  /cons_splice= (5'site:yes,3'site:no)
FT   variation      replace(3962,G)
FT                  /*tag= s
FT                  /standard_name= "single nucleotide polymorphism"
FT   variation      replace(4086,G)
FT                  /*tag= t
FT                  /standard_name= "single nucleotide polymorphism"
FT   exon           4088..4191
FT                  /*tag= d
FT                  /number= 2
FT   intron         4192..15055
FT                  /*tag= e
FT                  /number= 2
FT   variation      replace(4249,C)
FT                  /*tag= u
FT                  /standard_name= "single nucleotide polymorphism"
FT   variation      replace(13656..13657,TGG)
```

```
FT   variation       /tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(14719,A)
FT                   /tag= w
FT   exon            /standard_name= "Single nucleotide polymorphism"
FT                   15056..15210
FT                   /tag= f
FT                   /number= 3
FT   intron          15211..15308
FT                   /tag= g
FT                   /number= 3
FT                   /cons_splice= (5'site:no,3'site:no)
FT   exon            15309..15477
FT                   /tag= h
FT                   /number= 4
FT   intron          15478..50572
FT                   /tag= i
FT                   /number= 4
FT   variation       replace(15950..15952,TT)
FT                   /tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16390,T)
FT                   /tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16787,C)
FT                   /tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21582,G)
FT                   /tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23764,G)
FT                   /tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(49353,C)
FT                   /tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(49419,A)
FT                   /tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(50550,G)
FT                   /tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            50573..50696
FT                   /tag= j
FT                   /number= 5
FT   intron          50697..53614
FT                   /tag= k
FT                   /number= 5
FT   variation       replace(50738,C)
FT                   /tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            53615..53730
FT                   /tag= l
FT                   /number= 6
FT   intron          53731..68111
FT                   /tag= m
FT                   /number= 6
FT   variation       /cons_splice= (5'site:yes,3'site:no)
FT                   replace(59039,G)
FT                   /tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            68112..68326
FT                   /tag= n
FT                   /number= 7
FT   intron          68327..69735
FT                   /tag= o
FT                   /number= 7
FT   variation       /cons_splice= (5'site:yes,3'site:no)
FT                   replace(69067..69068,AMC)
FT                   /tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            69736..69844
FT                   /tag= p

FT   intron          /number= 8
FT                   69845..70296
FT                   /tag= q
FT   exon            /number= 8
FT                   70297..70543
FT                   /tag= r
FT   variation       /number= 9
FT                   replace(71085,G)
FT                   /tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   WO20027 9252-A1.
XX
PD   10-OCT-2002.
XX
PF   10-JAN-2002; 2002WO-US000456.
XX
PR   22-JAN-2001; 2001US-0262658P.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Merkulov G, Beasley EM;
XX
DR   WPI; 2003-040658/03.
DR   P-PSDB; ABG71742.
XX
PT   New isolated human transporter proteins, useful for developing
PT   therapeutic or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate transporter activity in cells or
PT   tissues.
XX
PS   Claim 4; Fig 3; 129pp; English.
XX
CC   The invention discloses an isolated human transporter polypeptide, and
CC   its allelic variants or orthologues, and the polynucleotides encoding
CC   them. Transporter proteins regulate many different functions of the cell,
CC   including cell proliferation, differentiation and signalling processes,
CC   by regulating the flow of molecules, such as ions and macromolecules,
CC   into and out of cells. The polynucleotide and polypeptide can be used to
CC   raise antibodies, create a gene chip, create a transgenic non-human
CC   animal, produce the novel polypeptide, detect the presence of the
CC   polypeptide or nucleic acid in a sample, identify a modulator of the
CC   polypeptide or its expression, identify a pharmaceutical composition, and
CC   carrier, that binds to the polypeptide and treat a disease or condition
CC   mediated by a human transporter protein which comprises administering to
CC   a patient the composition identified. The peptides and nucleic acid
CC   molecules are also useful in the development of human therapeutics and
CC   diagnostic compositions. The peptides are also useful for eliciting an
CC   immune response, to quantitatively determine levels of the protein (or
CC   its binding partner or ligand) in biological fluids or as markers for
CC   tissues in which the corresponding protein is preferentially expressed.
CC   The agents identified are useful for treating transporter-related
CC   conditions and a modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human
CC   transporter protein gene
XX
SQ   Sequence 73544 BP; 17350 A; 11721 C; 11526 G; 17753 T; 0 U; 15194 Other;
```

```
Alignment Scores:
Pred. No.:
Score:                    1.4e+04        Length:          73544
Percent Similarity:       9.00           Matches:         9
Best Local Similarity:    100.0%         Conservative:    0
Query Match:              100.0%         Mismatches:      0
DB:                       4.2%           Indels:          0
                          8              Gaps:            0

US-09-989-890-238 (1-212) x ABS6296 (1-73544)

Qy      24 LeuProSerSerLeuAlaLeuLeuSer 32
           |||||||||||||||||||||||||||
Db   70413 TTGCCATCCTCTCTGGCCTTGTTGTCC 70439
```

```
RESULT 97
ADC85298/c
ID   ADC85298 standard; DNA; 96599 BP.
XX
AC   ADC85298;
XX
DT   01-JAN-2004 (first entry)
XX
DE   Human Egr2 coding sequence.
XX
KW   Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW   secreted; transmembrane; intracellular; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003045230-A2.
XX
PD   05-JUN-2003.
XX
PF   02-DEC-2002; 2002WO-US038582.
XX
PR   30-NOV-2001; 2001US-00997722.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW, Engelhard EK;
XX
DR   WPI; 2003-513603/48.
XX
PT   New recombinant nucleic acid comprising a nucleotide sequence of any of
PT   the carcinoma-associated (CA) genes, useful for screening for drug
PT   candidates for diagnosing or treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 84; 983pp; English.
XX
CC   The invention relates to a recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the fully defined carcinoma-
CC   associated (CA) genes from the 50 tables given in the specification. The
CC   CA proteins are secreted, transmembrane or intracellular proteins. The
CC   recombinant nucleic acids are useful for screening for drug candidates
CC   for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC   ADC85514 represent CA genes of the invention.
XX
SQ   Sequence 96599 BP; 27390 A; 19000 C; 20559 G; 29350 T; 0 U; 300 Other;

Alignment Scores:
Pred. No.:                    1.8e+04     Length:        96599
Score:                        9.00        Matches:       9
Percent Similarity:           100.0%      Conservative:  0
Best Local Similarity:        100.0%      Mismatches:    0
Query Match:                  4.2%        Indels:        0
DB:                           10          Gaps:          0

US-09-989-890-238 (1-212) x ADC85298 (1-96599)

Qy      54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
           ||||||||||||||||||||||||||||||
Db   29167 CTCCCTAGTGCCAGTGCTGCCAGCTGGT 29141

RESULT 98
ADA02819/c
ID   ADA02819 standard; DNA; 96600 BP.
XX
AC   ADA02819;
XX
DT   06-NOV-2003 (first entry)
XX
DE   Mouse Sos1 carcinoma associated gene, SEQ ID NO:1337.
XX
KW   Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW   prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW   gene; ds.
XX
```

```
OS   Mus sp.
XX
PN   WO2003057146-A2.
XX
PD   17-JUL-2003.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 1337; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed murine CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 96600 BP; 27391 A; 19001 C; 20558 G; 29350 T; 0 U; 300 Other;

Alignment Scores:
Pred. No.:                    1.8e+04     Length:        96600
Score:                        9.00        Matches:       9
Percent Similarity:           100.0%      Conservative:  0
Best Local Similarity:        100.0%      Mismatches:    0
Query Match:                  4.2%        Indels:        0
DB:                           9           Gaps:          0

US-09-989-890-238 (1-212) x ADA02819 (1-96600)

Qy      54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
           ||||||||||||||||||||||||||||||
Db   29168 CTCCCTAGTGCCAGTGCTGCCAGCTGGT 29142

RESULT 99
ADB72557/c
ID   ADB72557 standard; DNA; 96600 BP.
XX
AC   ADB72557;
XX
DT   04-DEC-2003 (first entry)
XX
DE   Mouse Sos1 gene.
XX
KW   mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
```

Left column

OS   Mus sp.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW, Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 385; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a mouse gene of the invention.
XX
SQ   Sequence 96600 BP; 27391 A; 19001 C; 20558 G; 29350 T; 0 U; 300 Other;

Alignment Scores:
Pred. No.:   1.8e+04   Length: 96600
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 10   Gaps: 0

US-09-989-890-238 (1-212) x ADB72557 (1-96600)

Qy   54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
          ||||||||||||||||||||||||||||||
Db 29168 CTCCCTAGTGCCAGTGCTGCAGCTGGT 29142

RESULT 100
ADM74414/c
ID   ADM74414 standard; DNA; 96600 BP.
XX
AC   ADM74414;
XX
DT   01-JUL-2004 (first entry)
XX
DE   Murine carcinoma associated (CA) nucleic acid #43.
XX
KW   Mouse; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW   carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW   cytostatic.
XX
OS   Mus musculus.
XX
PN   US2004072154-A1.
XX
PD   15-APR-2004.
XX
PF   30-NOV-2001; 2001US-00997722.
XX
PR   22-DEC-2000; 2000US-00747377.
PR   02-MAR-2001; 2001US-00798586.

Right column

XX
PA   (MORR/) MORRIS D W.
PA   (ENGE/) ENGELHARD E K.
XX
PI   Morris DW, Engelhard EK;
XX
DR   WPI; 2004-328562/30.
XX
PT   New carcinoma associated gene or protein, useful for preparing a
PT   composition for diagnosing or treating carcinoma e.g., leukemia or
PT   lymphoma.
XX
PS   Claim 1; SEQ ID NO 85; 29pp; English.
XX
CC   The invention relates to new recombinant nucleic acids. The invention
CC   also relates to a host cell comprising a recombinant nucleic acid or
CC   expression vector, an expression vector comprising a recombinant nucleic
CC   acid, a recombinant protein, a method of screening for drug candidates, a
CC   method of screening for a bioactive agent capable of binding to a
CC   carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC   method of screening for a bioactive agent capable of modulating the
CC   activity of a CAP, a method of evaluating the effect of a candidate
CC   carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC   the activity of a CAP, a method of treating carcinoma, a method of
CC   neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC   propensity to carcinoma. A method of evaluating the effect of a candidate
CC   carcinoma drug comprises administering the drug to a patient, removing a
CC   cell sample from the patient and determining alterations in the
CC   expression or activation of a gene comprising the nucleotide sequence. A
CC   method of diagnosing carcinoma comprises determining the expression of
CC   one or more genes comprising the nucleic acid sequence in a first tissue
CC   type of a first individual and comparing the expression of the gene from
CC   a second normal tissue type from the first individual or a second
CC   unaffected individual, where a difference in the expression indicates
CC   that the first individual has carcinoma. A method of inhibiting the
CC   activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC   carcinomas comprises administering to a patient an inhibitor of CAP.
CC   Neutralising the effect of a CAP comprises contacting an agent specific
CC   for the CAP. The polypeptide specifically binds to the protein encoded by
CC   the nucleic acid. It comprises an antibody that specifically binds to the
CC   protein encoded by the nucleic acid. The nucleic acids are useful for
CC   preparing a composition for diagnosing or treating carcinoma e.g.,
CC   leukaemia or lymphoma. This sequence represents a murine carcinoma
CC   associated (CA) nucleic acid of the invention. Note: The sequence data
CC   for this patent did not form part of the printed specification but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 96600 BP; 27391 A; 19001 C; 20558 G; 29350 T; 0 U; 300 Other;

Alignment Scores:
Pred. No.:   1.8e+04   Length: 96600
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x ADM74414 (1-96600)

Qy   54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
          ||||||||||||||||||||||||||||||
Db 29168 CTCCCTAGTGCCAGTGCTGCAGCTGGT 29142

Search completed: March 17, 2006, 06:24:18
Job time : 793 secs

THIS PAGE LEFT BLANK

OM protein - nucleic search, using frame_plus_p2n model

                    GenCore version 5.1.7
          Copyright (c) 1993 - 2006 Biocceleration Ltd.

Run on: March 17, 2006, 06:12:02 ; Search time 3872 Seconds
                                        (without alignments)
                                        2561.689 Million cell updates/sec

Title: US-09-989-890-238
Perfect score: 212
Sequence: 1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table: OLIGO
                Xgapop 60.0 , Xgapext 60.0
                Ygapop 60.0 , Ygapext 60.0
                Fgapop   6.0 , Fgapext   7.0
                Delop   6.0 , Delext   7.0

Searched: 41078325 seqs, 2339341228 residues

Word size: 1

Total number of hits satisfying chosen parameters: 82152101

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095517_17553/app_query.fasta_1
-DB=EST -QFMT=fasta -SUFFIX=oligo_p2n.rst -MINMATCH=0.1 -LOOPCI=0 -LOOPEXT=0
-UNITS=bits -START=-1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=500
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=100 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss04
-USER=US09989890 @CGN_1_1_5315 @runat 16032006_095517_17553 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

                        SUMMARIES

Database : EST:*
            1: gb_est1:*
            2: gb_est2:*
            3: gb_est3:*
            4: gb_htc:*
            5: gb_est4:*
            6: gb_est5:*
            7: gb_est6:*
            8: gb_est7:*
            9: gb_gss1:*
           10: gb_gss2:*
           11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	212	100.0	1197	11	DQ045548	DQ045548 Homo sapi
2	177	83.5	917	5	BQ691555	BQ691555 AGENCOURT
3	150	70.8	966	2	BE901537	BE901537 AGENCOURT
4	148	69.8	510	1	AA316608	AA316608 EST188290
5	144	67.9	753	7	CN289134	CN289134 170005999
6	134	63.2	1671	6	CR749558	CR749558 Homo sapi
7	132	62.3	583	3	BP331524	BP331524 BP331524

No.	Score	Bits	Length	DB	ID	ID/Description
8		92	43.4	582	3	BP315176 BP315176
9		85	40.1	1015	3	BE741035 601594018
10		70	33.0	244	2	BE162756 PM1-HT045
11		58	27.4	1092	2	BE409561 601299865
12		51	24.1	1181	11	DQ045549 Pan trogl
13		48	22.6	623	2	BE741110 601593919
14		45	21.2	139	1	AW732798 bb14f07.y
15		44	20.8	975	2	BM460277 AGENCOURT
16		27	12.7	277	2	BQ081980 K-EST0026
17		20	9.4	276	7	CR763815 DKFZp469G
18		19	9.0	928	2	BG335647 602404274
19		19	9.0	1088	3	BG335025 602403470
20		12	5.7	425	2	BF150866 uy88g10.y
21		12	5.7	485	2	BF041373 BP250025B
22		12	5.7	609	2	BE533148 601235445
23		12	5.7	657	5	BY753126 BY753126
24		12	5.7	680	5	BI653517 603300520
25		12	5.7	685	5	BI156000 602903535
26		12	5.7	722	3	BI409054 602961119
27		12	5.7	738	2	BI078885 602873115
28		12	5.7	746	3	BI159562 602919525
29		12	5.7	746	3	BI905713 603166180
30		12	5.7	756	3	BG871324 602792482
31		12	5.7	944	3	BI905189 603167516
32		12	5.7	1040	3	BI411303 602964692
33	c	11	5.2	328	9	AQ418417 RPCI-11-1
34	c	10	4.7	257	2	BF115064 hr73h01.x
35	c	10	4.7	277	1	AW197658 xm85b08.x
36	c	10	4.7	296	2	BF590030 naa48h12.
37	c	10	4.7	297	2	AI382453 ta72f03.x
38	c	10	4.7	312	1	BY130650 BY130650
39	c	10	4.7	322	1	AL973065 AL973065
40	c	10	4.7	325	5	BY125377 BY125377
41	c	10	4.7	341	3	BP423898 BP423898
42	c	10	4.7	347	3	BY336928 BY336928
43	c	10	4.7	349	3	BY338512 BY338512
44	c	10	4.7	350	5	BY311584 BY311584
45	c	10	4.7	350	5	CD598944 RK137A3B1
46	c	10	4.7	352	5	BY343852 BY343852
47	c	10	4.7	355	5	BY338857 BY338857
48	c	10	4.7	362	2	BF733006 nae07d09.
49	c	10	4.7	365	1	AI631631 wb01c08.x
50	c	10	4.7	388	5	BY019540 BY019540
51	c	10	4.7	401	2	BB863674 UI-M-BH0-
52	c	10	4.7	404	6	CA949507 iq25c06.x
53	c	10	4.7	419	1	AI494178 t14g02.y
54	c	10	4.7	421	1	AI854714 UI-M-BH0-
55	c	10	4.7	434	1	AI188507 qd14e10.x
56	c	10	4.7	459	1	AW467069 ha09a02.x
57	c	10	4.7	460	1	AW420053 fj86g08.y
58	c	10	4.7	488	6	CA941290 ir33b08.x
59	c	10	4.7	504	1	AV780819 AV780819
60	c	10	4.7	511	5	BF434989 7p04c04.x
61	c	10	4.7	557	9	BH298354 CH230-146
62	c	10	4.7	560	7	CC740775 ZMMBb0010
63	c	10	4.7	576	7	CK691303 ZF101-P00
64	c	10	4.7	597	11	CR181960 Reverse s
65	c	10	4.7	600	11	AW300151 xe58b08.x
66	c	10	4.7	605	10	CE564205 tigr-gss-
67	c	10	4.7	618	5	BX871243 BX871243
68	c	10	4.7	620	2	BB659723 BB659723
69	c	10	4.7	647	2	BI824376 603038602
70	c	10	4.7	648	2	BB393441 BB393441
71	c	10	4.7	650	6	CD598528 RK114A2E1
72	c	10	4.7	652	2	BM265585 fw57d06.y
73	c	10	4.7	659	7	CR976136 CR976136
74	c	10	4.7	668	3	BM051939 603638841
75	c	10	4.7	686	5	BM051456 603638116
76	c	10	4.7	695	8	DR001418 TC118890
77	c	10	4.7	702	6	CF727607 UI-M-HB0-
78	c	10	4.7	705	6	BX890304 BX890304
79	c	10	4.7	710	6	CB246007 UI-M-FO0-
80	c	10	4.7	711	8	DR608449 EST998577

flag	idx			acc	acc			desc
c	81	10	4.7	DR614353	DR614353	714	8	EST100448
c	82	10	4.7	CC437175	CC437175	714	9	PUHBS30TB
c	83	10	4.7	BH353352	BH353352	721	5	CH230-68M
c	84	10	4.7	BX913421	BX913421	724	8	BX913421
	85	10	4.7	DR605627	DR605627	724	8	EST995755
	86	10	4.7	CC437178	CC437178	724	5	PUHBS30TD
c	87	10	4.7	BX912624	BX912624	726	5	BX912624
	88	10	4.7	CK688894	CK688894	733	7	ZF101-P00
c	89	10	4.7	CE265375	CE265375	744	10	tigr-gss-
c	90	10	4.7	CF386451	CF386451	749	6	RTDR1_14
	91	10	4.7	DR652062	DR652062	749	8	EST104217
	92	10	4.7	DR647955	DR647955	757	8	EST103807
c	93	10	4.7	CX773156	CX773156	758	3	UI-EH-HG0
	94	10	4.7	DR646867	DR646867	760	8	EST103698
	95	10	4.7	CK473171	CK473171	769	7	AGENCOURT
c	96	10	4.7	BQ180749	BQ180749	783	3	UI-M-EXO-
	97	10	4.7	CK478920	CK478920	783	7	AGENCOURT
c	98	10	4.7	DR656203	DR656203	785	8	EST104632
c	99	10	4.7	BZB12231	BZB12231	809	9	PUFCY39TD
c	100	10	4.7	BQ770347	BQ770347	810	5	UI-M-FIO-
	101	10	4.7	DR630229	DR630229	815	8	EST102035
	102	10	4.7	DN936585	DN936585	825	8	AGENCOURT
	103	10	4.7	DR659647	DR659647	827	8	EST104976
c	104	10	4.7	DR669323	DR669323	836	8	EST105944
c	105	10	4.7	CG177782	CG177782	836	10	PUKEA68TB
c	106	10	4.7	CNS04332	CNS04332	840	11	Tetraodon
c	107	10	4.7	DR654348	DR654348	846	8	EST104446
	108	10	4.7	DR629017	DR629017	847	8	EST101914
	109	10	4.7	CC404107	CC404107	847	9	PUHOL71TD
c	110	10	4.7	BI952193	BI952193	848	3	HVSMEm000
c	111	10	4.7	DR644266	DR644266	850	8	EST103489
	112	10	4.7	DR675398	DR675398	861	8	EST106551
c	113	10	4.7	CN127129	CN127129	862	3	RHOH1_21
c	114	10	4.7	DR667035	DR667035	862	8	EST105715
	115	10	4.7	DR638363	DR638363	867	8	EST102898
c	116	10	4.7	DR656474	DR656474	876	8	EST104659
c	117	10	4.7	DR669322	DR669322	888	8	EST105943
c	118	10	4.7	DR656473	DR656473	890	8	EST104659
	119	10	4.7	DR648421	DR648421	891	8	EST103853
c	120	10	4.7	CG455404	CG455404	892	10	PUIJV75TD
c	121	10	4.7	CG177783	CG177783	897	10	PUKEA68TD
c	122	10	4.7	DR675399	DR675399	903	8	EST106551
	123	10	4.7	DR656202	DR656202	910	8	EST104631
c	124	10	4.7	CC368485	CC368485	915	8	PUHRH49TB
c	125	10	4.7	DR659646	DR659646	917	8	EST104976
c	126	10	4.7	DR633270	DR633270	918	8	EST102389
	127	10	4.7	DR648422	DR648422	920	8	EST103853
c	128	10	4.7	BQ712803	BQ712803	928	5	AGENCOURT
	129	10	4.7	DR652964	DR652964	928	8	EST104308
	130	10	4.7	BQ941361	BQ941361	935	2	AGENCOURT
c	131	10	4.7	DR654347	DR654347	938	8	EST104446
	132	10	4.7	DR652063	DR652063	942	8	EST104218
c	133	10	4.7	DR640988	DR640988	949	8	EST103161
	134	10	4.7	DR652965	DR652965	954	8	EST104308
	135	10	4.7	DR636057	DR636057	959	8	EST102668
c	136	10	4.7	CC368490	CC368490	966	9	PUHRH49TD
	137	10	4.7	BQ433524	BQ433524	974	7	AGENCOURT
	138	10	4.7	CG455364	CG455364	1000	10	PUIJV75TB
c	139	10	4.7	CF618711	CF618711	1068	6	AGENCOURT
	140	10	4.7	AG055459	AG055459	1309	10	GA__Ea001
	141	10	4.7	AW727465	AW727465	1348	1	GA__Ea001
	142	9	4.2	AK043961	AK043961	1658	9	Mus muscu
c	143	9	4.2	AK048938	AK048938	2050	4	Mus muscu
	144	9	4.2	BH670120	BH670120	128	9	BOMMG74TR
c	145	9	4.2	AZ587920	AZ587920	149	9	1M0395K21
	146	9	4.2	BH891510	BH891510	190	9	3526_1_18
c	147	9	4.2	BZ329464	BZ329464	196	5	hv86a11.g
	148	9	4.2	BQ904602	BQ904602	205	9	Ta04_04h0
	149	9	4.2	BH229320	BH229320	216	9	1006151E0
	150	9	4.2	BH639879	BH639879	222	9	1008032E1
c	151	9	4.2	AI562046	AI562046	225	1	vs16f06.x
	152	9	4.2	BM871331	BM871331	231	3	mgns014xA
c	153	9	4.2	CN885048	CN885048	265	7	010622AAS

flag	idx					acc	acc	desc
	154	9	4.2	296	1	AA691613	AA691613	vs16f06.r
	155	9	4.2	296	6	CB639604	CB639604	OSJNBa11E
	156	9	4.2	296	1	BH229365	BH229365	1006151G0
	157	9	4.2	301	2	BB123886	BB123886	BB123886
	158	9	4.2	306	2	BB752317	BB752317	BB752317
c	159	9	4.2	308	10	BB664559	BB664559	152225 MA
c	160	9	4.2	311	2	CL173042	CL173042	104_376_1
	161	9	4.2	321	2	BI022859	BI022859	CM4-MT024
c	162	9	4.2	321	5	BQ767942	BQ767942	EBro08_SQ
	163	9	4.2	322	1	AA367730	AA367730	EST78881
	164	9	4.2	330	10	C2647476	C2647476	OM__Ba019
c	165	9	4.2	335	2	BF721665	BF721665	mab20g10.
c	166	9	4.2	348	1	AA208251	AA208251	mv82d07.r
	167	9	4.2	352	2	BF458818	BF458818	UI-M-BZ1-
	168	9	4.2	352	7	BY233516	BY233516	BY233516
c	169	9	4.2	352	1	CK743742	CK743742	wmi01-3ms
	170	9	4.2	362	1	AA326411	AA326411	EST29586
	171	9	4.2	367	3	BI022873	BI022873	CM4-MT024
	172	9	4.2	368	1	AA791451	AA791451	vs63c11.r
c	173	9	4.2	369	6	CF142207	CF142207	UI-HF-BR0
	174	9	4.2	370	3	BP828132	BP828132	BP828132
c	175	9	4.2	374	2	BF734813	BF734813	MR0-KT000
c	176	9	4.2	375	8	DN175444	DN175444	NMB05022
c	177	9	4.2	377	7	CO942580	CO942580	UMC-p2mm3
	178	9	4.2	379	2	BE955423	BE955423	UI-M-BH4-
	179	9	4.2	392	1	AV801750	AV801750	AV801750
c	180	9	4.2	393	10	AW352862	AW352862	35759 MAR
	181	9	4.2	393	3	CE594833	CE594833	tigr-gss-
	182	9	4.2	394	3	BJ215301	BJ215301	BJ215301
	183	9	4.2	395	5	BZ651064	BZ651064	OGAMA35TC
	184	9	4.2	397	5	BY362419	BY362419	BY362419
c	185	9	4.2	397	6	CD936365	CD936365	OV.104J22
c	186	9	4.2	398	1	AI757959	AI757959	EtESTea35
c	187	9	4.2	399	2	BF955895	BF955895	QV1-NN121
	188	9	4.2	400	5	BY613286	BY613286	BY613286
c	189	9	4.2	400	9	BZ651071	BZ651071	OGAMA35TM
c	190	9	4.2	402	1	AI325210	AI325210	ms29f04.x
c	191	9	4.2	404	2	AI759551	AI759551	EtESTea22
	192	9	4.2	405	1	AV817803	AV817803	AV817803
c	193	9	4.2	405	5	BY039243	BY039243	BY039243
c	194	9	4.2	406	2	BY430962	BY430962	BY430962
c	195	9	4.2	406	7	BE424990	BE424990	WHE0079_H
	196	9	4.2	409	1	CN885059	CN885059	010622AAS
c	197	9	4.2	410	2	AA613608	AA613608	no16b11.s
	198	9	4.2	410	1	AA140018	AA140018	mq39f04.r
	199	9	4.2	410	1	AV790154	AV790154	AV790154
c	200	9	4.2	411	1	AA725920	AA725920	vu85a08.r
	201	9	4.2	413	1	AQ122136	AQ122136	HS_3083_A
c	202	9	4.2	419	10	AG979127	AG979127	Drosophil
	203	9	4.2	422	2	BB731686	BB731686	BB731686
	204	9	4.2	422	10	BE627664	BE627664	uu53h02.y
c	205	9	4.2	422	2	CE833991	CE833991	tigr-gss-
c	206	9	4.2	424	7	CN945003	CN945003	011004AVB
c	207	9	4.2	425	1	AU194182	AU194182	AU194182
	208	9	4.2	426	10	CL969873	CL969873	OsIFCC019
c	209	9	4.2	442	5	BY652549	BY652549	BY652549
c	210	9	4.2	448	5	BY259166	BY259166	BY259166
	211	9	4.2	449	1	AV794074	AV794074	AV794074
	212	9	4.2	449	1	AW123857	AW123857	UI-M-BH2.
c	213	9	4.2	456	1	BJ215302	BJ215302	BJ215302
	214	9	4.2	457	9	AW763657	AW763657	ur63h09.x
	215	9	4.2	457	6	BH858862	BH858862	S3_x026a.
	216	9	4.2	457	10	CZ917795	CZ917795	4021007A0
	217	9	4.2	461	6	CD037482	CD037482	mgsu012xG
	218	9	4.2	461	10	CZ917888	CZ917888	4021007C1
	219	9	4.2	462	7	BE979415	BE979415	UI-M-BG2-
c	220	9	4.2	464	1	AW502073	AW502073	UI-HF-BR0
c	221	9	4.2	466	10	CL148948	CL148948	104_329_1
	222	9	4.2	467	2	BF913017	BF913017	MR3-UT010
c	223	9	4.2	468	3	BI786538	BI786538	sai50c10.
	224	9	4.2	474	1	BB693238	BB693238	BB693238
c	225	9	4.2	475	2	AI007181	AI007181	ua72g11.r
c	226	9	4.2	478	2	BE956450	BE956450	UI-M-BH4-

Left column:

#				acc	n	accession	description
227	c	9	4.2	BF913882	2	BF913882	MR3-UT010
228		9	4.2	BF913027	2	BF913027	MR3-UT010
229	c	9	4.2	BX614610	5	BX614610	BX614610
230	c	9	4.2	AI844656	1	AI844656	UI-M-AL1-
231		9	4.2	CD037483	9	CD037483	CD037483
232		9	4.2	AQ797168	9	AQ797168	nbxb0092P
233		9	4.2	AI510591	1	AI510591	mq39f04.y
234		9	4.2	CD037493	6	CD037493	mgsu014xM
235		9	4.2	CD037495	6	CD037495	mgsu014xO
236	c	9	4.2	BI400936	3	BI400936	MI-P-AY1-
237		9	4.2	CD037475	6	CD037475	mgsu011xI
238		9	4.2	BM370656	3	BM370656	BBro08_SQ
239		9	4.2	CD037479	6	CD037479	mgsu011xM
240	c	9	4.2	CD898494	6	CD898494	G174.109D
241	c	9	4.2	CA891399	3	CA891399	WHE0961_C
242	c	9	4.2	BE499543	6	BE499543	WHE0961_C
243	c	9	4.2	CD924140	6	CD924140	G750.11lL
244		9	4.2	AV435005	1	AV435005	AV435005
245	c	9	4.2	CN687776	7	CN687776	E0247C12-
246	c	9	4.2	BB758659	7	BB758659	BB758659
247		9	4.2	CN374552	7	CN374552	rzhswab0
248		9	4.2	AZ504750	9	AZ504750	1M0345O0I
249	c	9	4.2	AJ825442	9	AJ825442	AJ825442
250	c	9	4.2	CD901372	6	CD901372	G356.103J
251	c	9	4.2	CL801243	10	CL801243	OR_CBa001
252		9	4.2	BG509404	9	BG509404	sad1d03.
253	c	9	4.2	CL374662	10	CL374662	RPCI44_47
254		9	4.2	BM094486	3	BM094486	saj16e07.
255	c	9	4.2	CF142852	5	CF142852	UI-HF-BR0
256		9	4.2	BQ566109	5	BQ566109	gi52a07.y
257		9	4.2	CL757530	10	CL757530	OR_BBa012
258	c	9	4.2	CX201846	8	CX201846	MNS03340
259	c	9	4.2	BZ126307	9	BZ126307	CH230-405
260		9	4.2	BZ589485	9	BZ589485	3590_1_70
261	c	9	4.2	CA699269	9	CA699269	wlk8_pK00
262	c	9	4.2	BM085633	3	BM085633	saj27e02.
263		9	4.2	CE538411	10	CE538411	tigr-gss-
264	c	9	4.2	BM031662	2	BM031662	497131 MA
265		9	4.2	BG838946	2	BG838946	Gc02_01g0
266		9	4.2	BE571034	3	BE571034	601334548
267		9	4.2	BI975329	3	BI975329	483671 MA
268		9	4.2	AI670127	1	AI670127	we65g04.x
269	c	9	4.2	CE731015	10	CE731015	tigr-gss-
270	c	9	4.2	CF142346	6	CF142346	UI-HF-BR0
271		9	4.2	BZ589839	7	BZ589839	3590_1_72
272	c	9	4.2	CK518366	7	CK518366	rswjb0_00
273	c	9	4.2	BZ589866	9	BZ589866	3590_1_72
274	c	9	4.2	BJ243219	3	BJ243219	BJ243219
275	c	9	4.2	BP235232	3	BP235232	BP235232
276		9	4.2	BP352430	3	BP352430	BP352430
277	c	9	4.2	CD822895	6	CD822895	BN25_046P
278	c	9	4.2	BP280648	3	BP280648	BP280648
279		9	4.2	BZ589411	9	BZ589411	3590_1_69
280	c	9	4.2	CF872290	7	CF872290	tric00ixo
281	c	9	4.2	BJ247372	3	BJ247372	BJ247372
282	c	9	4.2	CL934388	10	CL934388	OA_ABa004
283	c	9	4.2	BJ208020	3	BJ208020	BJ208020
284		9	4.2	CB455208	6	CB455208	712337 MA
285		9	4.2	AL516999	1	AL516999	AL516999
286		9	4.2	AI064694	9	AI064694	HA0443 Hu
287		9	4.2	BH735681	9	BH735681	BOHTG58TR
288	c	9	4.2	BJ256339	9	BJ256339	BJ256339
289		9	4.2	CN883059	7	CN883059	010813AAS
290	c	9	4.2	CZ099065	10	CZ099065	OM_Ba010
291		9	4.2	BQ379706	5	BQ379706	MR3-UT010
292	c	9	4.2	BX299365	9	BX299365	BX299365
293		9	4.2	CA079113	6	CA079113	SCSBAM108
294		9	4.2	BZ589343	9	BZ589343	3590_1_69
295	c	9	4.2	CN868297	7	CN868297	010411AAO
296	c	9	4.2	AG981536	10	AG981536	Drosophil
297		9	4.2	BQ867487	5	BQ867487	QGD10P19.
298		9	4.2	BZ333036	9	BZ333036	hx32d03.b
299	c	9	4.2	BF484769	2	BF484769	WHE2319_D

Right column:

#					v	n	accession	description
300	c	9	4.2	627	7	CN881907	CN881907	010621AAS
301		9	4.2	630	6	CA286846	CA286846	SCSBSD205
302	c	9	4.2	630	9	CC801185	CC801185	ih11b03.b
303	c	9	4.2	634	8	DR041570	DR041570	BNE00019
304		9	4.2	636	9	CB454766	CB454766	711577 MA
305		9	4.2	642	4	BZ891837	BZ891837	Hm10_0106
306		9	4.2	644	9	BH545990	BH545990	BOHCJ89TR
307	c	9	4.2	645	4	AZ987727	AZ987727	2M0270G22
308		9	4.2	647	5	BX299366	BX299366	BX299366
309		9	4.2	651	4	BJ226656	BJ226656	BJ226656
310	c	9	4.2	655	8	DN400813	DN400813	LIB4004-0
311		9	4.2	658	3	BJ208021	BJ208021	BJ208021
312	c	9	4.2	660	6	AF524556	AF524556	AF524556
313		9	4.2	661	9	CD938085	CD938085	OV.109A10
314		9	4.2	663	5	BX627225	BX627225	BX627225
315	c	9	4.2	665	5	BY706420	BY706420	BY706420
316	c	9	4.2	668	3	AK006352	AK006352	Mus muscu
317	c	9	4.2	671	2	BB409186	BB409186	601300173
318	c	9	4.2	673	10	AG077325	AG077325	Pan trogl
319		9	4.2	674	1	AW208371	AW208371	uo59a01.x
320		9	4.2	675	3	BJ261872	BJ261872	BJ261872
321		9	4.2	676	3	BJ220053	BJ220053	BJ220053
322	c	9	4.2	677	10	CE603442	CE603442	tigr-gss-
323		9	4.2	682	10	BI959174	BI959174	HVSMEn001
324	c	9	4.2	684	6	CD924076	CD924076	G750.111G
325	c	9	4.2	684	7	CO889222	CO889222	BovGen_17
326		9	4.2	686	3	BJ249259	BJ249259	BJ249259
327		9	4.2	689	7	BJ231578	BJ231578	BJ231578
328		9	4.2	691	10	AG307369	AG307369	Mus muscu
329	c	9	4.2	695	10	CL807599	CL807599	OR_CBa002
330	c	9	4.2	698	7	CO040141	CO040141	UI-M-EMO-
331		9	4.2	700	6	CD938084	CD938084	OV.109A10
332		9	4.2	700	7	CO040030	CO040030	UI-M-EMO-
333		9	4.2	700	8	DN123650	DN123650	1123971 M
334		9	4.2	702	8	BB632760	BB632760	BB632760
335		9	4.2	702	6	CD904408	CD904408	G356.113I
336	c	9	4.2	706	9	BY752940	BY752940	BY752940
337		9	4.2	706	10	CW018674	CW018674	ZMMBLc001
338	c	9	4.2	709	8	CX207012	CX207012	MNS11575
339	c	9	4.2	710	8	DN047551	DN047551	ssallkgd0
340		9	4.2	713	5	BY731881	BY731881	BY731881
341		9	4.2	713	9	BH928245	BH928245	od182c10.
342	c	9	4.2	713	2	CZ087168	CZ087168	OM_Ba008
343		9	4.2	714	9	BG698956	BG698956	602703377
344	c	9	4.2	715	10	CW279864	CW279864	104_754_1
345		9	4.2	717	3	BJ253423	BJ253423	BJ253423
346		9	4.2	717	5	BQ870276	BQ870276	QGD8J17.y
347		9	4.2	718	6	CA227869	CA227869	SCQGFL305
348		9	4.2	720	5	CD898495	CD898495	G174.109D
349		9	4.2	720	8	CD901373	CD901373	G356.103J
350	c	9	4.2	729	6	DR657412	DR657412	EST104752
351	c	9	4.2	729	7	BZ657180	BZ657180	OGCCS05TC
352		9	4.2	730	10	CW493955	CW493955	Pan trogl
353		9	4.2	734	10	AG142114	AG142114	Pan trogl
354		9	4.2	743	4	BZ063523	BZ063523	lle86b03.
355		9	4.2	746	7	CO040519	CO040519	UI-M-EMO-
356		9	4.2	747	7	CO040286	CO040286	UI-M-EMO-
357		9	4.2	748	10	CE824826	CE824826	tigr-gss-
358		9	4.2	749	5	BQ870569	BQ870569	QGD9G17.y
359		9	4.2	751	2	BG567042	BG567042	602589360
360	c	9	4.2	754	10	AG430917	AG430917	Mus muscu
361	c	9	4.2	757	7	BI082647	BI082647	602878049
362	c	9	4.2	759	10	AG586453	AG586453	Mus muscu
363		9	4.2	760	9	DT063544	DT063544	AGENCORT
364		9	4.2	761	5	BX083350	BX083350	BX083350
365	c	9	4.2	763	10	AG603040	AG603040	Mus muscu
366	c	9	4.2	765	5	BX911162	BX911162	BX911162
367		9	4.2	766	2	BG934577	BG934577	SKI-0888
368	c	9	4.2	766	8	DR614087	DR614087	EST100421
369		9	4.2	766	9	DR619750	DR619750	EST100987
370	c	9	4.2	766	10	CZ326906	CZ326906	ZMMBF0032
371		9	4.2	767	6	CF067380	CF067380	Ac65 Amph
372		9	4.2	767	7	CK446113	CK446113	pncs914aA

idx				accession	description
373	9	4.2	9	CC595762	CH240_397
374	9	4.2	8	DR614498	EST100462
375	9	4.2	10	CG823768	SOYDT23TH
376	9	4.2	9	DR616735	EST100686
c 377	9	4.2	5	BX888940	BX888940
378	9	4.2	9	BZ657182	OGCCS05TM
379	9	4.2	9	DN123986	1124355 M
380	9	4.2	10	C2881611	C2881611_OC_Ba028
c 381	9	4.2	8	DR612065	EST100219
382	9	4.2	8	DR615082	EST100521
c 383	9	4.2	9	BH527515	BOGKT95TF
384	9	4.2	9	CJ021499	CJ021499
385	9	4.2	7	CA301524	SCUTSD208
386	9	4.2	6	CD750727	AGENCOURT
387	9	4.2	9	CC468079	CH240_139
c 388	9	4.2	10	CL301135	CL301135_gbs2_CH25
389	9	4.2	6	CA752750	mgae0005d
390	9	4.2	10	CZ354287	ZMMBF0091
391	9	4.2	8	DR621348	EST101147
392	9	4.2	7	DN108748	1106279 M
393	9	4.2	9	CK147436	AGENCOURT
394	9	4.2	7	CK774923	963726 MA
395	9	4.2	3	BI694617	603347838
396	9	4.2	4	BX054729	Single re
397	9	4.2	9	CNS09EE5	CNS09EE5
398	9	4.2	5	BZ165454	CH230-463
399	9	4.2	9	BU929170	AGENCOURT
400	9	4.2	4	BZ264227	CH230-326
401	9	4.2	6	CD377923	PTMM03092
402	9	4.2	10	C2337300	ZMMBF0048
403	9	4.2	10	AG902471	Oryza sat
404	9	4.2	10	CC411367	PUHRM91TB
405	9	4.2	7	CO798923	AGENCOURT
406	9	4.2	5	CD375796	PTMM00961
407	9	4.2	6	BG369040	HVSMEi002
c 408	9	4.2	2	BG918168	602821088
c 409	9	4.2	5	BG843520	1024003B0
c 410	9	4.2	6	CD384120	PTMM09292
411	9	4.2	1	AA203126	zx57g02.r
c 412	9	4.2	10	DN109102	1106663 M
413	9	4.2	2	CL512362	SAIL_859_
414	9	4.2	8	DR094329	STRR1_13
415	9	4.2	2	DR650481	EST104059
416	9	4.2	2	BF033546	601453508
417	9	4.2	9	BI249865	602995967
c 418	9	4.2	2	DR650480	EST104059
419	9	4.2	9	CC697321	OGUDV94TV
c 420	9	4.2	7	CL424434	AUF_IpSto
c 421	9	4.2	5	CL710892	OR_BBa003
c 422	9	4.2	10	C2856731	C2856731_OC_Ba024
c 423	9	4.2	10	AG890144	Oryza sat
424	9	4.2	10	AG850175	Oryza sat
c 425	9	4.2	7	CV070364	WPAEhux15
426	9	4.2	10	AG331916	Mus muscu
427	9	4.2	10	AG876239	Oryza sat
428	9	4.2	2	BF144315	601787134
c 429	9	4.2	10	AG168258	Pan trogl
430	9	4.2	10	CG106182	PUFNU84TB
431	9	4.2	2	BG547826	602576117
c 432	9	4.2	10	BG283331	602407061
433	9	4.2	2	CC921766	t05a a01ba
434	9	4.2	5	BU940763	AGENCOURT
435	9	4.2	8	DR911779	EST112331
436	9	4.2	2	BF135747	601781110
437	9	4.2	2	AZ684906	ENTLL05TR
438	9	4.2	9	BM926918	AGENCOURT
439	9	4.2	10	CL124979	ISB1-86E5
440	9	4.2	4	BX041901	Single re
c 441	9	4.2	2	CNS094HT	CNS094HT
c 442	9	4.2	10	AG070562	Pan trogl
443	9	4.2	5	BY707606	BY707606
c 444	9	4.2	11	CNS03R1A	CNS03R1A
c 445	9	4.2	8	DN972093	H05_ISUFS

idx				pos	accession	description
446	9	4.2	10	1022	CL468454	CL468454 SAIL_1287
447	9	4.2	9	1039	AQ900370	AQ900370 HS_2013_B
c 448	9	4.2	9	1050	BF037252	BF037252 601460963
c 449	9	4.2	8	1072	DR741268	DR741268 FGAS00119
450	9	4.2	6	1076	CF596057	CF596057 AGENCOURT
451	9	4.2	9	1084	BM455229	BM455229 AGENCOURT
452	9	4.2	10	1096	CL085974	CL085974 ISB1-6G14
c 453	9	4.2	10	1134	CG749697	CG749697 P044-1-A1
454	9	4.2	10	1142	CL478508	CL478508 SAIL_291
455	9	4.2	10	1154	AG063307	AG063307 Pan trogl
c 456	9	4.2	3	1162	BM805851	BM805851 AGENCOURT
457	9	4.2	9	1164	CC218420	CC218420 CH261-44M
c 458	9	4.2	3	1243	BM545338	BM545338 AGENCOURT
459	9	4.2	3	1266	BM803028	BM803028 AGENCOURT
c 460	9	4.2	3	1272	BC052051	BC052051 Mus muscu
461	9	4.2	3	1343	BM807280	BM807280 AGENCOURT
462	9	4.2	3	1387	BI855351	BI855351 603382414
463	9	4.2	3	1475	CNS0ACV0	CNS0ACV0 Arabidops
464	9	4.2	3	1514	BM909394	BM909394 AGENCOURT
465	9	4.2	10	1594	CL500870	CL500870 SAIL_68_D
466	9	4.2	3	1596	AK051860	AK051860 Mus muscu
c 467	9	4.2	3	1635	BM466090	BM466090 AGENCOURT
468	9	4.2	4	1657	AK007551	AK007551 Mus muscu
469	9	4.2	4	1681	BF694876	BF694876 602080875
470	9	4.2	4	1746	AK075909	AK075909 Mus muscu
471	9	4.2	2	1977	BG033596	BG033596 602301720
c 472	9	4.2	4	2223	AY401549	AY401549 Mus muscu
473	9	4.2	4	2246	AK078233	AK078233 Mus muscu
c 474	9	4.2	4	2498	BC027058	BC027058 Mus muscu
c 475	9	4.2	4	2923	AK077270	AK077270 Mus muscu
c 476	9	4.2	11	8830	DQ050467	DQ050467 Pan trogl
c 477	9	4.2	11	8899	DQ050466	DQ050466 Homo sapi
478	8	3.8	1	36	AW247788	AW247788 2820164.5
479	8	3.8	1	74	AL790515	AL790515 AL790515
480	8	3.8	4	83	AK213843	AK213843 Mus muscu
c 481	8	3.8	1	103	AI551335	AI551335 vx45c01.x
c 482	8	3.8	10	104	CG465711	CG465711 KRIBB_2D
c 483	8	3.8	2	111	BE184190	BE184190 CMO-HT067
c 484	8	3.8	6	115	CA785194	CA785194 sau25h10.
c 485	8	3.8	1	122	CNS08GWV	CNS08GWV Single re
486	8	3.8	10	123	CL273399	CL273399 Ggal_83a_
487	8	3.8	2	124	CX123098	CX123098 NMB01086
488	8	3.8	2	125	BI007209	BI007209 QV3-RT007
489	8	3.8	1	128	AI917403	AI917403 ts79b06.x
c 490	8	3.8	1	135	CX110268	CX110268 EI046G10
491	8	3.8	3	140	AV635303	AV635303 AV635303
c 492	8	3.8	7	140	CN724116	CN724116 E087D12-
c 493	8	3.8	7	144	BE685502	BE685502 187830 MA
494	8	3.8	7	145	CN710647	CN710647 E0580G09-
495	8	3.8	3	151	BP944835	BP944835 BP944835
496	8	3.8	1	152	CK495505	CK495505 rswbb0_00
497	8	3.8	1	155	AI147713	AI147713 qb43c09.x
c 498	8	3.8	8	157	DN889934	DN889934 nag29g03.
c 499	8	3.8	8	158	AF188509	AF188509 AF188509
c 500	8	3.8	7	159	CR530332	CR530332 CR530332

ALIGNMENTS

RESULT 1
DQ045548

```
LOCUS       DQ045548               1197 bp    DNA     linear   GSS 02-JUN-2005
DEFINITION  Homo sapiens FLJ34633 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   DQ045548
VERSION     DQ045548.1  GI:66896763
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1197)
```

```
AUTHORS     Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
            Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civello,D.,
            White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE       A Scan for Positively Selected Genes in the Genomes of Humans and
            Chimpanzees
JOURNAL     (er) PLoS Biol. 3 (6), E170 (2005)
PUBMED      15869325
REFERENCE   2 (bases 1 to 1197)
AUTHORS     Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
            Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civello,D.,
            White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE       Direct Submission
JOURNAL     Submitted (05-MAY-2005) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment. Translation starts at the beginning of
            alignment.
FEATURES             Location/Qualifiers
     source          1..1197
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
     gene            <1..>1197
                     /gene="FLJ34633"
                     /locus_tag="HC11301"
ORIGIN

Alignment Scores:
Pred. No.:               7.68e-182          Length:         1197
Score:                   212.00             Matches:        212
Percent Similarity:      100.0%             Conservative:   0
Best Local Similarity:   100.0%             Mismatches:     0
Query Match:             100.0%             Indels:         0
DB:                      11                 Gaps:           0

US-09-989-890-238 (1-212) x DQ045548 (1-1197)

Qy    1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104 AGCCCCCACCAAGCCGCCGCCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 163

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla      40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164 CAGAGAGCATTACCTTCATCTCTGCTCTTGCTGAGCCGCCCTTGAGTCCCCCACCTGCT   223

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGGCCGGGCTGCCTTCTGCTTCCGCCGCT 283

Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaCysLeu        80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  284 GCCGGGATTGCCTCCAGCCGTGGAGCCCTGTGTGCGGGAGATGCAGCCCCTGCCTGTCTA  343

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  344 CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC  403

Qy  101 ProProAlaLeuIleValHisProAlaGlyGlyMetAlaSerGlySerGlnPro      120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  404 CCCCCAGCCCTGATCGTGCACCCCGCAGGGGGATGGCCAGCGGCTCAAGTCAACCA    463

Qy  121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr   140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  464 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC 523

Qy  141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIle 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  524 CGAGGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGCTGCAGCCGCCAGGAGCCACTGGG 583

Qy  161 ProHisProCysAspThrAlaCysProLeuProLeuProValValLeuValAlaProArg 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  584 CCCCACCCATGCGACACAGCCTGCCCAGCACCTTTGCCAGTAGTCCTCGTGGCTCCGAGG 643
```

```
Qy  181 SerThrThrIleLeuSerMetSerArgTrpThrThrCysArgArgTrpProAlaValAlaProCys 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  644 AGTACTATTCTTCCATGAGTCGGACCTGGCCCGGAGATGGGCAGTGGCTCCAATGT          703

Qy  201 ArgAlaGluLysLeuMetCysSerSerSerArgSer 212
          ||||||||||||||||||||||||||||||||||||
Db  704 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 739

RESULT 2
BQ691555
LOCUS       BQ691555                917 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT 8341018 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6249453
            5', mRNA sequence.
ACCESSION   BQ691555
VERSION     BQ691555.1  GI:21816871
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 917)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2390 row: o column: 22
            High quality sequence stop: 535.
FEATURES             Location/Qualifiers
     source          1..917
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6249453"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site 2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:               3.88e-150          Length:         917
Score:                   177.00             Matches:        177
Percent Similarity:      100.0%             Conservative:   0
Best Local Similarity:   83.5%              Mismatches:     0
Query Match:             5                  Indels:         0
DB:                                         Gaps:           0

US-09-989-890-238 (1-212) x BQ691555 (1-917)

Qy    1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  147 AGCCCCCACCAAGCCGCCGCCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 206

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla      40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  207 CAGAGAGCATTACCTTCATCTCTGCTCTTGCTGAGCCGCCCTTGAGTCCCCCACCTGCT   266
```

ORIGIN

Alignment Scores:
Pred. No.:   1.42e-125   Length: 966
Score: 150.00   Matches: 189
Percent Similarity: 99.0%   Conservative: 0
Best Local Similarity: 99.0%   Mismatches: 2
Query Match: 70.8%   Indels: 2
DB: 2   Gaps: 0

US-09-989-890-238 (1-212) x BE901537 (1-966)

```
Qy    1  SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly   20
Db   10  AGCCCCCACCAAGCCGCCCGACCCCGTAGACACCAGACCCCAAGGACCCTGGC-ACCATGGGC   68

Qy   21  GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla   40
Db   69  CAGAGAGCATTACCTTCATCTCTGCTCTGCTCCGAGCGGCCCCTTGAGTCCCCCACCTGCT  128

Qy   41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla   60
Db  129  GCCTGCTCTCGGCGACCCTGGGGTGTCGGGAGTGGTGCCGGGCTGCCTTCGTCCGCCGCT  188

Qy   61  AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu   80
Db  189  GCCGGGGATTGCCTCGTCGGCGCTGTGGAGCCTGTGTGCGGGGATGCAGCCCCTGCCTGTCTA  248

Qy   81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
Db  249  CTGAGGGACTCCACTGAGGGAGTGCTGAAGCCAACTGGGCTCAAGGAGCACAATGAGTGC  308

Qy  101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro  120
Db  309  CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAGCGGCTCAAGTCAACCA  368

Qy  121  TrpProAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
Db  369  TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC  428

Qy  141  ArgGlyProProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIle  160
Db  429  CGAGGGGCCACCTCCCCAGCCCTGATGCGGCACCTCTGCTGCAAGGAGCCACTGGCCGATC  488

Qy  161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProArg  180
Db  489  CCCCACCCCATGCGACACAGC-TGCCCAGCACCTTTGCCAGTAGTCCTCGTGGCTCCGAGG  547

Qy  181  SerThrIleLeuSerMetSerArgThrTrpThr  191
Db  548  AGTACTATTCTTTCCATGAGTCGGACCTGACT  580
```

RESULT 4
AA316608
LOCUS   AA316608   510 bp   mRNA   linear   EST 19-APR-1997
DEFINITION EST188290 HCC cell line (matastasis to liver in mouse) II Homo
            sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA316608
VERSION   AA316608.1   GI:1968936
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 510)
AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,

---

```
Qy   41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla   60
Db  267  GCCTGCTCTCGGCGACCCCTGGGGTGTCGGGAGTGGTGCCGGGCTGCCTTCGTCCGCCGCT  326

Qy   61  AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu   80
Db  327  GCCGGGATTGCCTCCAGCCGTGGAGCCTGTGTGCGGGGATGCAGCCCCTGCCTGTCTA  386

Qy   81  LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys  100
Db  387  CTGAGGGACTCCACTGAGGGAGTGCTGAAGCCAACTGGCCAAGGAGCACAATGGAGTGC  446

Qy  101  ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerGlnPro  120
Db  447  CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAGCGGCTCAAGTCAACCA  506

Qy  121  TrpProAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr  140
Db  507  TGGGCAGCAGCTTCAGCTACCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTATCCCTACC  566

Qy  141  ArgGlyProProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIle  160
Db  567  CGAGGGGCCACCTCCCCAGCCCTGATGCGGACTCCTGTGCTGCAAGGAGCCACTGGCCGATC  626

Qy  161  ProHisProCysAspThrAlaCysProAlaProLeuProValValLeuVal  177
Db  627  CCCCACCCCATGCGACACAGCACCTGCCCCAGCACCTTTGCCAGTAGTCCTCGTG  677
```

RESULT 3
BE901537
LOCUS   BE901537   966 bp   mRNA   linear   EST 29-SEP-2000
DEFINITION 601677420F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3960220 5',
            mRNA sequence.
ACCESSION   BE901537
VERSION   BE901537.1   GI:10390818
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 966)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM842 row: g column: 05
            High quality sequence stop: 668.
FEATURES   Location/Qualifiers
     source   1..966
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3960220"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_21"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."

```
Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
```

TITLE   Initial assessment of human gene diversity and expression patterns
           based upon 83 million nucleotides of cDNA sequence
JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
PUBMED   7566098
COMMENT   Contact: Kerlavage, AR
           Bioinformatics
           The Institute for Genomic Research
           9712 Medical Center Drive, Rockville, MD 20850 USA
           Tel: 3018699056
           Fax: 3018699423
           Email: arkerlav@tigr.org
           For clone availability, additional sequence and expression
           information related to this EST, please check the TIGR Human Gene
           Index (http://www.tigr.org/tdb/hgi/hgi.html)
           Seq primer: M13 Reverse.

FEATURES   Location/Qualifiers
     source   1..510
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="ATCC (inhost):113362"
                 /db_xref="taxon:9606"
                 /tissue_type="colon"
                 /cell_type="KM12SM"
                 /cell_line="KM12C(HCC)metastasis into mouse (liver)"
                 /cell_line="KM12C(HCC) cell line (metastasis to liver in mouse)
                 I"
                 /note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                 XhoI"

ORIGIN

Alignment Scores:
Pred. No.:   Length: 510
Score: 5.1e-124   Matches: 148
Percent Similarity: 148.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 69.8%   Gaps: 0
                                                       1

US-09-989-890-238 (1-212) x AA316608 (1-510)

Qy 65 SerSerAlaValGluProValCysGlyAspAlaAlaProAlaAlaCysLeuLeuArgThrPro 84
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 TCCAGCGCTGTGGAGCCTGTGTGCGGGGATGCAGCCCCTGCTGCTTGCCTACTGAGGACTCCA 61

Qy 85 LeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCysProProAlaLeu 104
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 62 CTGAGGGGACTGCTGAAGCCAACTGGCCCAAGGAGCACATGGAGTGCCCCCCAGCCCTG 121

Qy 105 IleValHisProProAlaGlyGlyMetAlaGlySerGlnSerSerGlnTrpAlaAlaAla 124
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122 ATCGTGCACCCCCCAGCCGGCGGGGGATGGCAGCCCAGCCGGCTCAAGCATGGGACCAGCT 181

Qy 125 SerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThrGlyProPro 144
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 182 TCAGCTACCCCGATGTTAAGCTCAAAGGAGCCATCCTGTGTATCCCTACCCGAGGGCCACCT 241

Qy 145 ProGlnProLeuMetArgThrArgProAlaAlaAlaArgSerHisTrpProIleProHisProCys 164
       |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 242 CCCCAGCCCCTGATGCGGACTCGTGCAAGGAGCCACTGGCCGATCCCCACCCATGC 301

Qy 165 AspThrAlaCysProAlaProLeuProValValLeuValAlaProAlaArgSerThrIleLeu 184
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 302 GACACAGCCTGCCCAGCACCTTTGCCAGTAGTCCTGTGGCTCGAGGAGTACTATTCTT 361
```

```
Qy 185 SerMetSerArgThrTrpTrpThrCysArgArgTrpAlaAlaValAlaProCysArgAlaGluLys 204
       |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 362 TCCATGAGTCGGACCTGGACCTGGACGTGGCGAGATGGGAGTGGCTCCATGTGCAGCCCGAGAAA 421

Qy 205 LeuMetCysSerSerArgSer 212
       |||||||||||||||||
Db 422 TTGATGTGCTCATCTTCAAGAAGC 445
```

RESULT 5
CN289134
LOCUS   CN289134   753 bp   mRNA   linear   EST 16-MAY-2004
DEFINITION 1700059993935 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN289134
VERSION   CN289134.1 GI:47305548
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 753)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE   Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
   PUBMED   15146197
COMMENT   Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 753   Std Error: 0.00.

FEATURES   Location/Qualifiers
     source   1..753
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /tissue_type="embryonic stem cells, DMSO-treated H9 cell
                 line"
                 /clone_lib="GRN_PREHEP"
                 /note="oligo dT primed, full-length enriched cDNA library
                 from DMSO-treated hES cell line H9 (p22) maintained in
                 feeder-free conditions"

ORIGIN

Alignment Scores:
Pred. No.:   Length: 753
Score: 3.18e-120   Matches: 144.00
Percent Similarity: 144.00   Conservative: 157
Best Local Similarity: 99.4%   Mismatches: 0
Query Match: 99.4%   Indels: 1
DB: 67.9%   Gaps: 0

US-09-989-890-238 (1-212) x CN289134 (1-753)

```
Qy  56 SerAlaSerAlaAlaAlaGlyIleAlaSerAlaValGluProValCysGlyAspAla  75
       |||||||||||||||||||||||||||||||||||||||||||||||||
Db  11 TCTGCTTCCGCCGCTGCCGGGATTGCCTCCGCCGTCGTGTGGGACCTGTGTGGGGGATGCA  70

Qy  76 AlaProAlaCysLeuLeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArg  95
       |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71 GCCCCTGCCTGTCTACTGAGGAGCACTGAGGGGACTGCTGAAGCCAACTGGCCCAAGG 130

Qy  96 SerThrMetGluCysProProAla-LeuIleValHisProProAlaGlyGlyMetAlaSe 115
       ||||||||||||||||||||||||  |||||||||||||||||||||||||
Db 131 AGCACAATGGAGTGCCCCCCAGCTGCTGATCGTGCACCCCAGCCGCGGGATGGCCAG 190

Qy 115 rGlySerSerGlnProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSe 135
```

Alignment Scores:
Pred. No.:   Length: 1671
Score: 8.29e-111   Matches: 134
Percent Similarity: 134.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 63.2%   Gaps: 0
                          4

US-09-989-890-238 (1-212) x CR749558 (1-1671)

```
Qy     79  CysLeuLeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMet  98
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    288  TGTCTACTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGCCAAGGAGCACAATG  347

Qy     99  GluCysProProAlaLeuIleValHisProProAlaGlyMetAlaSerGlySerSer    118
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    348  GAGTGCCCCCCAGCCCTGATCGTGCACCCCCCAGCGGGCGGGGATGGCCAGCGGCTCAAGT  407

Qy    119  GlnProTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIle  138
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    408  CAACCATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATC  467

Qy    139  ProThrArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrp    158
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    468  CCTACCGAGGGCCACCTCCCCAGCCCCTGATGCGGACACCTTCTGCTGCAAGGAGCCACTGG  527

Qy    159  ProIleProHisProCysAspThrAlaCysProAlaProLeuProValLeuValLeuValAla  178
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    528  CCGATCCCCCACCCCATGCGACACAGCCTGCCCAGCACCTTTGCCAGTAGTCCTGTGGCT  587

Qy    179  ProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAla  198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    588  CCGAGGAGTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGCAGTGGCT  647

Qy    199  ProCysArgAlaGluLysLeuMetCysSerSerArgSer  212
           ||||||||||||||||||||||||||||||||||||||||
Db    648  CCATGTCGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC  689
```

RESULT 7
BP331524
LOCUS   BP331524   583 bp   mRNA   linear   EST 17-SEP-2004
DEFINITION   BP331524 Sugano cDNA library, rectum Homo sapiens cDNA clone
             RCT09504, mRNA sequence.
ACCESSION   BP331524
VERSION   BP331524.1 GI:52260875
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE   1 (bases 1 to 583)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
             Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE   Sequence comparison of human and mouse genes reveals a homologous
             block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
   PUBMED   15342556
COMMENT   Contact: Yutaka Suzuki
             Department of Virology
             Institute of Medical Science, University of Tokyo
             4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
             Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES   Location/Qualifiers
     source   1..583
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="RCT09504"
                     /tissue_type="rectum"
                     /clone_lib="Sugano cDNA library, rectum"
ORIGIN

```
Db    191  CGGCTCAAGTCAACCATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATC  250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    135  rLeuCysIleProThrArgGlyProProGlnProLeuMetArgThrProAlaAlaAr  155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251  CCTGTGTATCCCTACCGAGGGCCACCTCCCCAGCCCCTGATGCGGACACCTTCTGCTGCAAG  310

Qy    155  gSerHisTrpProIleProHisProCysAspThrAlaCysProAlaProLeuProValVa  175
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311  GAGCCACTGGCCGATCCCCCACCCCATGCGACACAGCCTGCCCAGCACCTTTGCCAGTAGT  370

Qy    175  lLeuValAlaProArgSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTr  195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371  CCTGTGGCTCCGAGGAGTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCGGAGATG  430

Qy    195  pAlaValAlaProCysArgAlaGluLysLeuMetCysSerSerArgSer  212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    431  GGCAGTGGCTCCATGTCGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC  482
```

RESULT 6
CR749558
LOCUS   CR749558   1671 bp   mRNA   linear   HTC 19-AUG-2004
DEFINITION   Homo sapiens mRNA; cDNA DKFZp686B17277 (from clone DKFZp686B17277).
ACCESSION   CR749558
VERSION   CR749558.1 GI:51476784
KEYWORDS   HTC.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE   1 (bases 1 to 1671)
  AUTHORS   Pousska,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
             Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
             Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE   Direct Submission
  JOURNAL   Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
             Neuherberg, GERMANY
COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
             sequenced by DKFZ (German Cancer Research Center,
             Heidelberg/Germany) within the cDNA sequencing consortium of the
             German Genome Project. This clone (DKFZp686B17277) is available at
             the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in
             Berlin, Germany. Please contact RZPD for ordering;
             http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686B17277
             Further information about the clone and the sequencing project is
             available at http://mips.gsf.de/projects/cdna/.
FEATURES   Location/Qualifiers
     source   1..1671
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686B17277"
                     /tissue_type="prostate"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
                     /note="hypothetical protein, differentially spliced"
     gene   1..1671
                     /gene="DKFZp686B17277"
     CDS   69..692
                     /gene="DKFZp686B17277"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH18355.1"
                     /db_xref="GI:51476785"
                     /translation="MPRPGHPRPASGPPRLGPWERPTELCLETYDKPPQPPSRRTRR
                     PDPKDPGHHGPESITFISGSAEPALESPTCLLRTPLRGLLKPTGPRSTMECPPALIVH
                     PPAGGMASGSSOPWAAASATPMLSSKASLCIPTRGPPPQPLMRTPAARSHWPIPHPCD
                     TACPAPLPVVLVAPRSTILSMSRTWTCRRWAVAPCRAEKLMCSSSRS"
ORIGIN
```

ORIGIN   /note="mammary gland tumor"

Alignment Scores:
Pred. No.:   4.59e-73   Length: 582
Score: Matches: 92
Percent Similarity: 92.00%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 43.4%   Gaps: 0
                          3

US-09-989-890-238 (1-212) x BP315176 (1-582)

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTGTATCCCTACC 65

Qy   141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaSerHisTrpProIle 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 CGAGGGCCACCTCCCCAGCCCGATGCGACTCCTGCTGCAAGGAGCCACTGGCCGATC 125

Qy   161 ProHisProCysAspThrAlaCysProAlaProLeuProValLeuValAlaProArg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 CCCCACCCATGGCGACACAGCCGTCCCAGCACTTTGCCAGTAGTCCTCCGTGGCTCCGAGG 185

Qy   181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   186 AGTACTATTCTTTCCATGAGTCGGACCTGGACCTGCCGGAGATGGGCAGTGGCTCCATGT 245

Qy   201 ArgAlaGluLysMetCysSerSerArgSer 212
           ||||||||||||||||||||||||||||||
Db   246 CGAGCCGAGAAATTGATGTGCTCATCTTCAAGAGC 281

RESULT 9
BB741035
LOCUS   BB741035   1015 bp   mRNA   linear   EST 15-SEP-2000
DEFINITION   601594018F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3947861 5',
             mRNA sequence.
ACCESSION   BB741035
VERSION   BB741035.1 GI:1015027
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE   1 (bases 1 to 1015)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM810 row: d column: 06
             High quality sequence stop: 752.
FEATURES   Location/Qualifiers
     source   1..1015
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3947861"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_9"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average

---

Alignment Scores:
Pred. No.:   2.02e-109   Length: 583
Score: 132.00   Matches: 132
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 62.3%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x BP331524 (1-583)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87 AGCCCCCACCAAGCCGCCGCACCCGTAGACCAGGACCAGACCCCGACCCTGGCCACCATGGGC 146

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   147 CAGAGAGCATTACCTTCATCTCTGCTCTGAGCCGGCCCTTGAGTCCCCACCTGCT 206

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyValAlaGlyLeuProSerAlaSerAlaAla 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   207 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTTCTTGCTTCCGGCCGCT 266

Qy   61 AlaGlyIleAlaSerArgAlaValAlaGlyProValCysGlyAspAlaAlaProAlaCysLeu 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   267 GCCGGGATTGCCTCCAGCCGTGTGGAGCCTGTGGGGGGATGCAGCCCCTGCCCTGTCTA 326

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   327 CTGAGGACTCCACTGAGGGGACTCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC 386

Qy   101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlySerSerGlnPro 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   387 CCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAAGCGGCTCAAGTCAACCA 446

Qy   121 TrpAlaAlaSerAlaThrProMetLeuSerSer 132
           ||||||||||||||||||||||||||||||||
Db   447 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCA 482

RESULT 8
BP315176
LOCUS   BP315176   582 bp   mRNA   linear   EST 17-SEP-2004
DEFINITION   BP315176 Sugano cDNA library, mammary gland OCUB-F Homo sapiens
             cDNA clone OFR07979, mRNA sequence.
ACCESSION   BP315176
VERSION   BP315176.1 GI:52244151
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE   1 (bases 1 to 582)
AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
             Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE   Sequence comparison of human and mouse genes reveals a homologous
             block structure in the promoter regions
JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
PUBMED   15342556
COMMENT   Contact: Yutaka Suzuki
             Department of Virology
             Institute of Medical Science, University of Tokyo
             4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
             Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES   Location/Qualifiers
     source   1..582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="OFR07979"
                     /tissue_type="mammary gland"
                     /cell_line="OCUB-F"
                     /clone_lib="Sugano cDNA library, mammary gland OCUB-F"

O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUMBED   10737800
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=PM1-HT0454-170
            100-003-f07&t3=2000-01-17&t4=1)
            Seg primer: puc 18 forward
            High quality sequence stop: 244.
FEATURES   Location/Qualifiers
     source   1..244
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0454"
                     /note="Organ: head neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 244
Score: 2.01e-53   Matches: 70
Percent Similarity: 70.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 33.0%   Gaps: 0

US-09-989-890-238 (1-212) x BE162756 (1-244)

Qy   56 SerAlaSerAlaAlaAlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAla   75
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 TCTGCTTCCGCCGCTGCCGGGATTGCCTCCAGCGCTGTGGAGCCTGTGTGCGGGGATGCA   90

Qy   76 AlaProAlaCysLeuLeuArgThrProLeuLeuLysProThrProGlyProArg   95
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91 GCCCCTGCCTGTCTACTGAGGACTCCACTGAGGGGACTCCTGAAGCCAACTGGGCCAAGG 150

Qy   96 SerThrMetGluCysProProAlaGluIleValHisProAlaGlyMetAlaSer   115
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151 AGCACAATGGAGTGCCCCCCAGCCCCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAGC 210

Qy   116 GlySerSerGlnProTrpAlaAlaAlaSer   125
            ||||||||||||||||||||||||||
Db   211 GGCTCCAAGTCAACCATGGGCAGCAGCTTCA   240

RESULT 11
BE409561
LOCUS   BE409561   1092 bp   mRNA   linear   EST 21-JUL-2000
DEFINITION 601299865F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3629848 5',
            mRNA sequence.
ACCESSION   BE409561
VERSION   BE409561.1 GI:9346011
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

---

insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 1015
Score: 1.79e-66   Matches: 191
Percent Similarity: 85.00   Conservative: 0
Best Local Similarity: 97.4%   Mismatches: 4
Query Match: 97.4%   Indels: 5
DB: 40.1%   Gaps: 2

US-09-989-890-238 (1-212) x BE741035 (1-1015)

Qy   18 ThrMetGlyGlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerPro   37
            |||||||||||||||||||||||||||||||||||||||||| |||||
Db   99 ACCATGGGCCAGAGAGCATTACCTTCATCTCTGGCTCTGAGCCTCGGCC--TTGAGTCCC 157

Qy   38 ProProAlaAlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAla   57
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   158 CCACCTGCTGCCTGCTCTGGCGACCCTGGGTGTGGGAGGTGCCGGGCTGCCTTCTGCT 217

Qy   58 SerAlaAlaAlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaPro   77
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   218 TCCGCCGCTGCCGGGATTGCCTCCAGCGCTGTGGAGCCTGTGTGCGGGGATGCAGCCCCT 277

Qy   78 AlaCysLeuLeuArgThr-ProLeuArgGlyLeuLeuLysProThrGlyProArgSerTh   97
            |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   278 GCCTGTCTACTGAGGAT-TCCACTGAGGGGACTCCTGAAGCCAACTGGGCCAAGGAGCAC 336

Qy   97 rMetGluCysProProAlaLeuIleValHisProAlaGlyMetAlaSerGlySe   117
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   337 AATGGAGTGCCCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAGCGGCTC 396

Qy   117 rSerGlnProTrpAlaAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCy 137
            |||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   397 AAGTCAACCATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCCTGTG 456

Qy   137 sIleProThrArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHi   157
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   457 TATCCCTACCCGGAGGGCCACCTCCCCAGCCCTTGATGCGGACTCCTGCTGCAAGGAGCCA 516

Qy   157 sTrpProIleProHisProCysAspThrAlaCysProAlaProLeuProValValLeuVa 177
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   517 CTGGCGGATCCCCACCCATGCGACACACCGCTG--CCAGCACCTTTGCCCAGTAGTCCTCGT 575

Qy   177 lSerGlnProHisProCysArgGlyMetSerSerArgThrTrpThrCysArgArgTrpAlaVa 197
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   576 GGCTCCAGGAGTACTATCTTTCCATGAGTCGGACCTGGACCTGCCCGAGATGGGCAGG 635

Qy   197 lAlaProCysArgAlaGluLysMetCysSerSerSerArgSer 212
            |||||||||||||||||||||||||||||||||||||||||
Db   636 -GCTCCATGTGCGAGCCGGAGAAATTGATGTGCTCATCTTCAAGAGAGC 680

RESULT 10
BE162756
LOCUS   BE162756   244 bp   mRNA   linear   EST 21-JUN-2000
DEFINITION PM1-HT0454-170100-003-F07 HT0454 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE162756
VERSION   BE162756.1 GI:8625477
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 244)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,

```
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 1181)
  AUTHORS   Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
            Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civello,D.,
            White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
  TITLE     A Scan for Positively Selected Genes in the Genomes of Humans and
            Chimpanzees
  JOURNAL   (er) PLoS Biol. 3 (6), E170 (2005)
   PUBMED   15869325
REFERENCE   2  (bases 1 to 1181)
  AUTHORS   Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B.,
            Hubisz,M.J., Fledel-Alon,A., Tanenbaum,D.M., Civello,D.,
            White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2005) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.  Translation starts at the beginning of
            alignment.
FEATURES             Location/Qualifiers
     source          1..1181
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     <1..>1181
     gene            /gene="FLJ34633"
                     /locus_tag="HC11301"
ORIGIN

Alignment Scores:
Pred. No.:                                 1181
Score:                     1.65e-35        Matches:                  51
Percent Similarity:        51.00           Conservative:              0
Best Local Similarity:     100.0%          Mismatches:                0
Query Match:               100.0%          Indels:                    0
DB:                        24.1%           Gaps:                      0
                           11

US-09-989-890-238 (1-212) x DQ045549 (1-1181)

Qy    1 SerProHisGlnAlaAlaAlaProValAspGlnThrArgThrLeuAlaThrMetGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104 AGCCCCCACCAAGCCGCCGGCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 163

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProProAla 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164 CAGAGGAGCATTACCTTCATCTCGGCTCTGCTGCTCCTGTCGAGTCCCCCACCTGCT 223

Qy   41 AlaCysSerGlyAspProGlyProGlyCysGlySerGly 51
          |||||||||||||||||||||||||||
Db  224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGN 256
```

```
RESULT 13
BB741110
LOCUS       BB741110                 623 bp    mRNA       linear   EST 15-SEP-2000
DEFINITION  601593919F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3947885 5',
            mRNA sequence.
ACCESSION   BB741110
VERSION     BB741110.1  GI:10155102
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
```

```
REFERENCE
  AUTHORS   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
  TITLE     NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM317 row: a column: 17
            High quality sequence start: 19
            High quality sequence stop: 482.
FEATURES             Location/Qualifiers
     source          1..1092
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3629848"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_21"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:                                 1092
Score:                     6.66e-42        Matches:                  84
Percent Similarity:        58.00           Conservative:              0
Best Local Similarity:     97.7%           Mismatches:                0
Query Match:               97.7%           Indels:                    2
DB:                        27.4%           Gaps:                      0
                           2

US-09-989-890-238 (1-212) x BB409561 (1-1092)

Qy    1 SerProHisGlnAlaAlaAlaProValAspGlnThrArgThrLeuAlaThrMetGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  208 AGCCCCCACCAAGCCGCCGGCACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 267

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeu-SerProProProAl 40
          ||||||||||||||||||||||||||||||||||||||||
Db  268 CAGAGGAGCATTACCTTCATCTCGGCTCTGCTGCTCCGAGCCGGCCCTTGAAGTCCCCCACCTGC 327

Qy   40 aAlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAl 60
          ||||||||||||||||||||||||||||||||||||||||
Db  328 TGCCTGCTCTGGCGACCCCTGGGTGGGAGTGGTGCCGGGGC 387

Qy   60 aAlaGlyIleAlaSerSerAlaValGlu-ProValCysGlyCysGlyAspAlaAlaProAlaCysL 80
          ||||||||||||||||||||||||||||||||||||||||
Db  388 TGCCGGGATTGCCTCCAGCGCGTGTGGGAGGCCTGTGCGGGGATGCAGCCCCCGCCTGTC 447

Qy   80 euLeuArgThrPro 84
          ||||||||||
Db  448 TACTGAGGACTCCC 461
```

```
RESULT 12
DQ045549
LOCUS       DQ045549                1181 bp    DNA        linear   GSS 02-JUN-2005
DEFINITION  Pan troglodytes FLJ34633 gene, VIRTUAL TRANSCRIPT, partial
            sequence, genomic survey sequence.
ACCESSION   DQ045549
VERSION     DQ045549.1  GI:66896764
```

COMMENT   DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -40RP from Gibco
High quality sequence stop: 112.

FEATURES   Location/Qualifiers
source   1..139
       /organism="Homo sapiens"
       /mol_type="mRNA"
       /db_xref="taxon:9606"
       /clone="IMAGE:2962885"
       /tissue_type="choriocarcinoma"
       /lab_host="DH10B (phage-resistant)"
       /clone_lib="NIH_MGC_21"
       /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:
Pred. No.:   6.24e-31   Length: 139
Score: 45.00   Matches: 45
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 21.2%   Indels: 0
DB: 1   Gaps: 0

US-09-989-890-238 (1-212) x AW732798 (1-139)

Qy   93 GlyProArgSerThrMetGluCysProProAlaLeuIleValHisProProAlaGlyGly 112
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3 GGGCCAAGGAGCACAATGGAGTGCCCCCCAGCCTGATCGTGCACCCCCCAGCCGGCGGG 62

Qy   113 MetAlaSerGlySerSerGlnProTrpArgAlaAlaSerAlaThrProMetLeuSerSer 132
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 ATGGCCAGCGGCTCAAGTCAACCATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCA 122

Qy   133 LysAlaSerLeuCys 137
       |||||||||||||||
Db   123 AAGGCATCCCTGTGT 137

RESULT 15
BM460277
LOCUS   BM460277   975 bp   mRNA   linear   EST 05-FEB-2002
DEFINITION   AGENCOURT_6420313 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5502726
5', mRNA sequence.
ACCESSION   BM460277
VERSION   BM460277.1   GI:18509317
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE   1 (bases 1 to 975)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished (1999)
JOURNAL
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM810 row: e column: 06
High quality sequence stop: 615.

FEATURES   Location/Qualifiers
source   1..623
       /organism="Homo sapiens"
       /mol_type="mRNA"
       /db_xref="taxon:9606"
       /clone="IMAGE:3947885"
       /tissue_type="adenocarcinoma cell line"
       /lab_host="DH10B (phage-resistant)"
       /clone_lib="NIH_MGC_9"
       /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:
Pred. No.:   4.82e-33   Length: 623
Score: 48.00   Matches: 48
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 22.6%   Indels: 0
                          Gaps: 2

US-09-989-890-238 (1-212) x BE741110 (1-623)

Qy   12 ThrProArgThrLeuAlaThrMetGlyLeuAsnArgAlaAlaLeuProSerSerLeuAlaLeuLeu 31
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   86 ACCCCAAGGACCCCTGGCCACCATGGGCCAGAGAGCATTACCTTACTTCATCTCTGGCTTCTGCTG 145

Qy   32 SerArgProLeuSerProProAlaAlaCysSerGlyAspProGlyCysGlySerGly 51
       ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   146 AGCCGGCCCTTGAGTCCCCCACCTGGCTGCCTGCCTCTGGCGACCCTGGGTGTGGGAGTTGGT 205

Qy   52 AlaGlyLeuProSerAlaSerAla 59
       ||||||||||||||||||||||||
Db   206 GCCGGGGCTGCCCTTCCGCTTCCGCC 229

RESULT 14
AW732798
LOCUS   AW732798   139 bp   mRNA   linear   EST 21-APR-2000
DEFINITION   bb14f07.y1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:2962885 5',
mRNA sequence.
ACCESSION   AW732798
VERSION   AW732798.1   GI:7633136
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE   1 (bases 1 to 139)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished (1999)
JOURNAL
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```
FEATURES
     source          Plate: LLAM12141 row: f column: 07
                     High quality sequence stop: 659.
                     Location/Qualifiers
                     1..975
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5502726"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Library constructed by Life
                     Technologies."
ORIGIN

Alignment Scores:
Pred. No.:
Score:                   3.18e-29      Length:          975
Percent Similarity:      44.00%        Matches:         44
Best Local Similarity:   100.0%        Conservative:    0
Query Match:             100.0%        Mismatches:      0
DB:                      20.8%         Indels:          0
                         3             Gaps:            0

US-09-989-890-238 (1-212) x BM460277 (1-975)

Qy  169  ProAlaProLeuProValValLeuValAlaProArgSerThrIleLeuSerMetSerArg 188
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2  CCAGCACCTTTGCCAGTAGTCCTCGTGGCTCCGAGGAGTACTATTCTTTCCATGAGTCGG 61

Qy  189  ThrTrpThrCysArgArgTrpAlaValAlaProCysArgAlaGluLysLeuMetCysSer 208
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62  ACCTGGACCTGCCGGAGATGGGCAGTGGCTCCATGTCGAGCCGAGAAATTGATGTGCTCA 121

Qy  209  SerSerArgSer 212
         ||||||||||||
Db  122  TCTTCAAGAGC 133

RESULT 16
BQ081980
LOCUS       BQ081980         277 bp    mRNA     linear    EST 04-APR-2002
DEFINITION  K-EST0026481 S9SNU601 Homo sapiens cDNA clone S9SNU601-20-C05 5',
            mRNA sequence.
ACCESSION   BQ081980
VERSION     BQ081980.1  GI:19938986
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 277)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.B., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 20 row: C column: 05
            High quality sequence stop: 277.
FEATURES             Location/Qualifiers
     source          1..277
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="S9SNU601-20-C05"
                     /sex="M"
                     /tissue_type="Ascites"
                     /cell_type="Epithelial"
                     /cell_line="SNU-601"
                     /lab_host="Top10F'"
                     /clone_lib="S9SNU601"
                     /note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI;
                     Site_2: XhoI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tobacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including SfiI
                     site by treatment of T4 RNA ligase and the first strand
                     cDNA was synthesized with Superscript II using SfiI
                     oligo-dT primer. After first strand synthesis, RNA was
                     degraded by NaOH treatment and cDNA was amplified by PCR
                     reaction. The PCR products were digested with SfiI and
                     cloned into DraIII- digested pME18S-FL3 vector. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library."
ORIGIN

Alignment Scores:
Pred. No.:                                              277
Score:                   2.75e-14      Length:          277
Percent Similarity:      27.00         Matches:         27
Best Local Similarity:   100.0%        Conservative:    0
Query Match:             100.0%        Mismatches:      0
DB:                      12.7%         Indels:          0
                         3             Gaps:            0

US-09-989-890-238 (1-212) x BQ081980 (1-277)

Qy   15  ThrLeuAlaThrMetGlyGlnArgArgAlaLeuProSerSerLeuAlaLeuSerArgPro 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  197  ACCCTGGCCACCATGGGCCAGAGGAGCATTACCTTCATCTCTTGCTCTGCTCGAGCCGGCC 256

Qy   35  LeuSerProProProAlaAla 41
         |||||||||||||||||||||
Db  257  TTGAGTCCCCCACCTGCTGCC 277

RESULT 17
CR763815
LOCUS       CR763815               276 bp    mRNA    linear    EST 23-SEP-2004
DEFINITION  DKFZp469G0138_r1 469 (synonym: pkid1) Pongo pygmaeus cDNA clone
            DKFZp469G0138 5', mRNA sequence.
ACCESSION   CR763815
VERSION     CR763815.1  GI:52602277
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pongo.
REFERENCE   1  (bases 1 to 276)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Ansorge,W., Krieger,S., Regiert,T., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert. Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; rlin, Germany. Please contact
            RZPD for ordering;
            http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp469G0138
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1..276
                     /organism="Pongo pygmaeus"
```

```
                                                          Score:                      19.00          Matches:        19
                                                          Percent Similarity:        100.0%          Conservative:    0
                                                          Best Local Similarity:     100.0%          Mismatches:      0
                                                          Query Match:                 9.0%          Indels:          0
                                                          DB:                             2           Gaps:           0

                                          US-09-989-890-238 (1-212) x BG335647 (1-928)

                                          Qy      16 LeuAlaThrMetGlyGlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgPro 34
                                                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                          Db     244 CTGGCCACCATGGGCCAGAGAGCATTACCTTCATCTCTGGCTCTCGAGCGGCCT 300

                                          RESULT 19
                                          BG335025
                                          LOCUS       BG335025                    1088 bp    mRNA      linear   EST 27-FEB-2001
                                          DEFINITION  602403470F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4541320 5',
                                                      mRNA sequence.
                                          ACCESSION   BG335025
                                          VERSION     BG335025.1  GI:13141463
                                          KEYWORDS    EST.
                                          SOURCE      Homo sapiens (human)
                                          ORGANISM    Homo sapiens
                                                      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                                                      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                                                      Hominidae; Homo.
                                          REFERENCE   1  (bases 1 to 1088)
                                          AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
                                          TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
                                          JOURNAL     Unpublished (1999)
                                          COMMENT     Contact: Robert Strausberg, Ph.D.
                                                      Email: cgapbs-r@mail.nih.gov
                                                      Tissue Procurement: ATCC
                                                      cDNA Library Preparation: Ling Hong/Rubin Laboratory
                                                      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                                                      DNA Sequencing by: Incyte Genomics, Inc.
                                                      Clone distribution: MGC clone distribution information can be
                                                      found through the I.M.A.G.E. Consortium/LLNL at:
                                                      http://image.llnl.gov
                                                      Plate: LLCM1219 row: k column: 17
                                                      High quality sequence stop: 556.
                                          FEATURES             Location/Qualifiers
                                                      source          1..1088
                                                                      /organism="Homo sapiens"
                                                                      /mol_type="mRNA"
                                                                      /db_xref="taxon:9606"
                                                                      /clone="IMAGE:4541320"
                                                                      /tissue_type="choriocarcinoma"
                                                                      /lab_host="DH10B (phage-resistant)"
                                                                      /clone_lib="NIH_MGC_21"
                                                                      /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                                                                      Site_2: EcoRI; cDNA made by oligo-dT priming.
                                                                      Directionally cloned into EcoRI/XhoI sites using the
                                                                      following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                                                                      for average insert size 1.8kb. Library constructed by
                                                                      Ling Hong in the laboratory of Gerald M. Rubin (University
                                                                      of California, Berkeley) using ZAP-cDNA synthesis kit
                                                                      (Stratagene) and Superscript II RT (Life Technologies)."
                                          ORIGIN

                                          Alignment Scores:
                                          Pred. No.:                1088
                                          Score:                 1.86e-06          Length:         1088
                                          Percent Similarity:      19.00           Matches:          19
                                          Best Local Similarity:  100.0%           Conservative:      0
                                          Query Match:            100.0%           Mismatches:        0
                                          DB:                       9.0%           Indels:            0
                                                                      2            Gaps:              0

                                          US-09-989-890-238 (1-212) x BG335025 (1-1088)

                                          Qy      18 ThrMetGlyGlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSer 36
                                                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                          Db     286 ACCATGGGCCAGAGAGCATTACCTTCATCTCTGGCTCTCGAGCGGCCCTTGAGT 342
```

```
/mol_type="mRNA"
/db_xref="taxon:9600"
/clone="DKFZp469G0138"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="469 (synonym: pkid1)"
/note="Vector: pSport1_Sfi; Site_1: SfiIa; Site_2: SfiIB"

ORIGIN

Alignment Scores:
Pred. No.:
Score:                 6.31e-08          Length:          276
Percent Similarity:      20.00           Matches:          20
Best Local Similarity:  100.0%           Conservative:      0
Query Match:            100.0%           Mismatches:        0
DB:                       9.4%           Indels:            0
                            7            Gaps:              0

US-09-989-890-238 (1-212) x CR763815 (1-276)

Qy      18 ThrMetGlyGlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     217 ACCATGGGCCAGAGAGCATTACCTTCATCTCTGGCTCTCGAGCGGCCCCCTGAGTCCC 276

RESULT 18
BG335647
LOCUS       BG335647                     928 bp    mRNA      linear   EST 27-FEB-2001
DEFINITION  602404274F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4541898 5',
            mRNA sequence.
ACCESSION   BG335647
VERSION     BG335647.1  GI:13142085
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 928)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1221 row: c column: 19
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
            source          1..928
                            /organism="Homo sapiens"
                            /mol_type="mRNA"
                            /db_xref="taxon:9606"
                            /clone="IMAGE:4541898"
                            /tissue_type="choriocarcinoma"
                            /lab_host="DH10B (phage-resistant)"
                            /clone_lib="NIH_MGC_21"
                            /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                            Site_2: EcoRI; cDNA made by oligo-dT priming.
                            Directionally cloned into EcoRI/XhoI sites using the
                            following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                            for average insert size 1.8kb. Library constructed by
                            Ling Hong in the laboratory of Gerald M. Rubin (University
                            of California, Berkeley) using ZAP-cDNA synthesis kit
                            (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:              1.6e-06          Length:          928
```

RESULT 20
BF150866
LOCUS   BF150866   uy88g10.y1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3666690 5',   425 bp   mRNA   linear   EST 29-DEC-2000
DEFINITION   uy88g10.y1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3666690 5',
               mRNA sequence.
ACCESSION   BF150866
VERSION   BF150866.1   GI:11032261
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 425)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
               Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Other ESTs: uy88g10.x1
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Washington University Genome Sequencing Center
               Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               image.llnl.gov/image/html/iresources.shtml

MGI:1427458
Seq primer: -40RP from Gibco
High quality sequence stop: 395.
FEATURES   Location/Qualifiers
     source   1..425
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="mix FVB/N, C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3666690"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam5"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"

ORIGIN

Alignment Scores:
Pred. No.:   1.77   Length: 425
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.7%   Indels: 0
DB: 2   Gaps: 0

US-09-989-890-238 (1-212) x BF150866 (1-425)

Qy   53 GlyLeuProSerAlaSerAlaSerAlaAlaAlaGlyIleAla 64
        ||||||||||||||||||||||||||||||||||||||||||||
Db   81 GGGCTGCCTTCGCTTCCGCCGCTGCAGGGATTGCC 116

RESULT 21
BF041373
LOCUS   BF041373   BP250025B20B12 Soares normalized bovine placenta Bos taurus cDNA   485 bp   mRNA   linear   EST 10-OCT-2000
DEFINITION   BP250025B20B12 Soares normalized bovine placenta Bos taurus cDNA
               clone BP250025B20B12 5', mRNA sequence.
ACCESSION   BF041373
VERSION   BF041373.1   GI:10758428
KEYWORDS   EST.
SOURCE   Bos taurus (cow)

ORGANISM   Bos taurus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
               Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 485)
AUTHORS   Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and
               Larson,J.H.
TITLE   Bovine ESTs
JOURNAL   Unpublished (2000)
COMMENT   Contact: Lewin, H. A.
               W. M. Keck Center for Comparative and Functional Genomics
               University of Illinois at Urbana-Champaign
               340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
               61801, USA
               Tel: 217 333 5998
               Fax: 217 244 5617
               Email: h-lewin@uiuc.edu
               Funding for cattle EST sequencing was provided by the USDA National
               Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
               to H. A. Lewin and J. B. Womack. Base Calling/Quality Scores: PHRED
               from Washington University Genome Center. Vector Trimmi g:
               Cross match from Washington University Genome Center PHRAP suite.
               This sequence is vector free and at least 200 bp in length.
               PCR PRimers
               FORWARD: TAATACGACTCACTATAGGG
               BACKWARD: ATTAACCCTCACTAAAG
               Insert Length: 485   Std Error: 0.00
               Plate: BP250025B20   row: B column: 12
               Seq primer: AGCGGATAACAATTTCACACAGGA
               High quality sequence stop: 485.

FEATURES   Location/Qualifiers
     source   1..485
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BP250025B20B12"
                     /sex="female"
                     /lab_host="DH10B"
                     /clone_lib="Soares normalized bovine placenta"
                     /note="Organ: placenta; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI; The cDNA library was contributed by the
                     Soares laboratory and it was constructed and normalized
                     as described by Bonaldo, M.F., Lennon, G. and Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. "

ORIGIN

Alignment Scores:
Pred. No.:   2   Length: 485
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.7%   Indels: 0
DB: 2   Gaps: 0

US-09-989-890-238 (1-212) x BF041373 (1-485)

Qy   36 SerProProAlaAlaCysSerGlyAspProGly 47
        |||||||||||||||||||||||||||||||||
Db 269 AGCCCCCCGCCTGCTGCCTGCTCTGGCGACCCTGGG 304

RESULT 22
BE533148
LOCUS   BE533148   601235445F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3599314 5',   609 bp   mRNA   linear   EST 09-AUG-2000
DEFINITION   601235445F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3599314 5',
               mRNA sequence.
ACCESSION   BE533148
VERSION   BE533148.1   GI:9761793
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE   1 (bases 1 to 609)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM8781 row: 1 column: 11
             High quality sequence stop: 587.
FEATURES   Location/Qualifiers
     source   1..609
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3599314"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
ORIGIN

US-09-989-890-238 (1-212) x BE533148 (1-609)

Alignment Scores:
Pred. No.:   2.48   Length: 609
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.7%   Indels: 0
DB: Gaps: 0

Qy   53 GlyLeuProSerAlaSerAlaAlaAlaAlaGlyIleAla 64
            |||||||||||||||||||||||||||||||||||||||
Db   290 GGGCTGCCTTCGCTTCCGCCGCCGCTGCAGGGATTGCC 325

RESULT 23
LOCUS   BY753126   657 bp   mRNA   linear   EST 17-DEC-2002
DEFINITION BY753126 RIKEN full-length enriched, adult inner ear Mus musculus
            cDNA clone F930110008 5', mRNA sequence.
ACCESSION   BY753126
VERSION   BY753126.1   GI:27184397
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 657)
AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
             Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
             Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
             Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
             Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
             Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
             Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
             Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
             Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
             Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
             Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
             Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,

Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
PUBMED   12466851
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC) Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome-res@gsc.riken.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
            Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
            Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
            Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
            Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
            Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
            Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Kirk W. Beisel (Boys Town National
            Research Hospital 555 North 30th Street Omaha,NE 68131 USA) whose
            assistance we gratefully acknowledge.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..657
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="F930110008"
                     /tissue_type="inner ear"
                     /dev_stage="adult"
                     /clone_lib="RIKEN full-length enriched, adult inner ear"
ORIGIN

Alignment Scores:
Pred. No.:   2.67   Length: 657
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0

Query Match: 5.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x BY753126 (1-657)

Qy   53 GlyLeuProSerAlaSerAlaAlaAlaGlyIleAla 64
            |||||||||||||||||||||||||||||||||||||
Db   524 GGGCTGCCTTCTGCTTCCGCCGCTGCAGGGATTGCC 559

RESULT 24
BI653517
LOCUS   BI653517   680 bp   mRNA   linear   EST 12-SEP-2001
DEFINITION   603300520F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5341085 5',
              mRNA sequence.
ACCESSION   BI653517
VERSION   BI653517.1 GI:15567753
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 680)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM1866 row: g column: 06
              High quality sequence stop: 676.
FEATURES   Location/Qualifiers
     source   1..680
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5341085"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam3"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
ORIGIN

Alignment Scores:
Pred. No.:   2.76   Length: 680
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.7%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x BI653517 (1-680)

Qy   53 GlyLeuProSerAlaSerAlaAlaAlaGlyIleAla 64
            |||||||||||||||||||||||||||||||||||||
Db   310 GGGCTGCCTTCTGCTTCCGCCGCTGCAGGGATTGCC 345

RESULT 25
BI156000
LOCUS   BI156000   685 bp   mRNA   linear   EST 05-JUL-2001

DEFINITION   602903535F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5033097 5',
              mRNA sequence.
ACCESSION   BI156000
VERSION   BI156000.1 GI:14616001
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 685)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM1092 row: f column: 10
              High quality sequence stop: 685.
FEATURES   Location/Qualifiers
     source   1..685
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5033097"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam3"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
ORIGIN

Alignment Scores:
Pred. No.:   2.78   Length: 685
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.7%   Indels: 0
DB: 2   Gaps: 0

US-09-989-890-238 (1-212) x BI156000 (1-685)

Qy   53 GlyLeuProSerAlaSerAlaAlaAlaGlyIleAla 64
            |||||||||||||||||||||||||||||||||||||
Db   524 GGGCTGCCTTCTGCTTCCGCCGCTGCAGGGATTGCC 559

RESULT 26
BI409054
LOCUS   BI409054   722 bp   mRNA   linear   EST 14-AUG-2001
DEFINITION   602961119F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5116780 5',
              mRNA sequence.
ACCESSION   BI409054
VERSION   BI409054.1 GI:15169977
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 722)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.

TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
             Bonaldo, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM11284 row: e column: 05
             High quality sequence stop: 720.

FEATURES   Location/Qualifiers
     source   1..722
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Czech II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5116780"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGACGGCGGCCCTCTGTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

US-09-989-890-238 (1-212) x BI409054 (1-722)

Alignment Scores:
Pred. No.:   2.92   Length: 722
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.7%   Indels: 0
DB: Gaps: 3

Qy   53 GlyLeuProSerAlaSerAlaAlaAlaAlaGlyIleAla 64
          |||||||||||||||||||||||||||||||||||||||
Db   440 GGGCTGCCTTCTGCTTCCGCCGCTGCAGGGATTGCC 475

RESULT 27
BI078885
LOCUS   BI078885   738 bp   mRNA   linear   EST 20-JUN-2001
DEFINITION   602873115F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5005226 5',
             mRNA sequence.
ACCESSION   BI078885
VERSION   BI078885.1 GI:14497215
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 738)
AUTHORS   NIH-MGC http://www.mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM11044 row: m column: 03
             High quality sequence stop: 735.

FEATURES   Location/Qualifiers
     source   1..738
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5005226"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam2"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"

ORIGIN

Alignment Scores:
Pred. No.:   2.98   Length: 738
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.7%   Indels: 0
DB: Gaps: 2

US-09-989-890-238 (1-212) x BI078885 (1-738)

Qy   53 GlyLeuProSerAlaSerAlaAlaAlaAlaGlyIleAla 64
          |||||||||||||||||||||||||||||||||||||||
Db   493 GGGCTGCCTTCTGCTTCCGCCGCTGCAGGGATTGCC 528

RESULT 28
BI159562
LOCUS   BI159562   746 bp   mRNA   linear   EST 05-JUL-2001
DEFINITION   602919525F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5059854 5',
             mRNA sequence.
ACCESSION   BI159562
VERSION   BI159562.1 GI:14619563
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 746)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM1162 row: a column: 07
             High quality sequence start: 3
             High quality sequence stop: 742.

FEATURES   Location/Qualifiers
     source   1..746
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5059854"
                     /tissue_type="tumor, gross tissue"

went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Alignment Scores:
```
Pred. No.:                3.01        Length:            746
Score:                    12.00       Matches:            12
Percent Similarity:       100.0%      Conservative:        0
Best Local Similarity:    100.0%      Mismatches:          0
Query Match:              5.7%        Indels:              0
DB:                                   Gaps:                0
```

US-09-989-890-238 (1-212) x BI905713 (1-746)

```
Qy   53 GlyLeuProSerAlaSerAlaAlaAlaGlyIleAla 64
        ||||||||||||||||||||||||||||||||||||
Db  420 GGGCTGCCTTCTGCTTCCGCCGCTGCAGGGATTGCC 455
```

RESULT 30
BG871324
LOCUS   BG871324   756 bp   mRNA   linear   EST 29-MAY-2001
DEFINITION   602792482F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4923597 5',
               mRNA sequence.
ACCESSION   BG871324
VERSION   BG871324.1   GI:14221864
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 756)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Jeffrey E. Green, M.D.
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM1845 row: c column: 22
               High quality sequence start: 114
               High quality sequence stop: 750.
FEATURES   Location/Qualifiers
     source   1..756
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4923597"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_SG2"
                     /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
                     NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
                     dT. Average insert size 1.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Alignment Scores:
```
Pred. No.:                3.05        Length:            756
Score:                    12.00       Matches:            12
Percent Similarity:       100.0%      Conservative:        0
Best Local Similarity:    100.0%      Mismatches:          0
Query Match:              5.7%        Indels:              0
DB:                                   Gaps:                2
```

US-09-989-890-238 (1-212) x BG871324 (1-756)

```
Qy   53 GlyLeuProSerAlaSerAlaAlaAlaGlyIleAla 64
```

/dev_stage="10 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam3"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."

ORIGIN

Alignment Scores:
```
Pred. No.:                3.01        Length:            746
Score:                    12.00       Matches:            12
Percent Similarity:       100.0%      Conservative:        0
Best Local Similarity:    100.0%      Mismatches:          0
Query Match:              5.7%        Indels:              0
DB:                                   Gaps:                0
```

US-09-989-890-238 (1-212) x BI159562 (1-746)

```
Qy   53 GlyLeuProSerAlaSerAlaAlaAlaGlyIleAla 64
        ||||||||||||||||||||||||||||||||||||
Db  517 GGGCTGCCTTCTGCTTCCGCCGCTGCAGGGATTGCC 552
```

RESULT 29
BI905713
LOCUS   BI905713   746 bp   mRNA   linear   EST 16-OCT-2001
DEFINITION   603166180F1 NCI_CGAP_Lu33 Mus musculus, cDNA clone IMAGE:5254507 5',
               mRNA sequence.
ACCESSION   BI905713
VERSION   BI905713.1   GI:16168280
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 746)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Gilbert Smith, Ph.D.
               cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
               Bonaldo, Ph.D.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM1642 row: o column: 20
               High quality sequence start: 3
               High quality sequence stop: 746.
FEATURES   Location/Qualifiers
     source   1..746
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Czech II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5254507"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGAGACGGAGCGGCCGCCCTCGTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library

Db   333 GGGCTGCCTTCTCGTTCCGCCGCTGCAGGCATTGCC 368
              |||||||||||||||||||||||||||||||||||||

RESULT 31
BI905189
LOCUS   BI905189   944 bp   mRNA   linear   EST 16-OCT-2001
DEFINITION   603167516F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5255666 5',
               mRNA sequence.
ACCESSION   BI905189
VERSION   BI905189.1 GI:16167680
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 944)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Gilbert Smith, Ph.D.
               cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
               Bonaldo, Ph.D.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by:Incyte Genomics, Inc.
               Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM1645 row: p column: 03
               High quality sequence start: 68
               High quality sequence stop: 927.
FEATURES   Location/Qualifiers
     source   1..944
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Czech II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5255666"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGAGCGGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN
Alignment Scores:
Pred. No.:   3.76   Length: 944
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x BI905189 (1-944)

Qy   53 GlyLeuProSerAlaSerAlaAlaAlaGlyIleAla 64
            |||||||||||||||||||||||||||||||||||||
Db   548 GGGCTGCCTTCTCGTTCCGCCGCTGCAGGGATTGCC 583

RESULT 32
BI411303
LOCUS   BI411303   1040 bp   mRNA   linear   EST 14-AUG-2001
DEFINITION   602964692F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5119942 5',

---

              mRNA sequence.
ACCESSION   BI411303
VERSION   BI411303.1 GI:15172226
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1040)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Gilbert Smith, Ph.D.
               cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
               Bonaldo, Ph.D.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by:Incyte Genomics, Inc.
               Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM1292 row: h column: 23
               High quality sequence start: 29
               High quality sequence stop: 909.
FEATURES   Location/Qualifiers
     source   1..1040
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Czech II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5119942"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGAGCGGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN
Alignment Scores:
Pred. No.:   4.12   Length: 1040
Score: 12.00   Matches: 12
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 5.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x BI411303 (1-1040)

Qy   53 GlyLeuProSerAlaSerAlaAlaAlaGlyIleAla 64
            |||||||||||||||||||||||||||||||||||||
Db   502 GGGCTGCCTTCTCGTTCCGCCGCTGCAGGGATTGCC 537

RESULT 33
AQ418417
LOCUS   AQ418417   328 bp   DNA   linear   GSS 23-MAR-1999
DEFINITION   RPCI-11-177D15.TJ RPCI-11 Homo sapiens genomic clone
               RPCI-11-177D15, genomic survey sequence.
ACCESSION   AQ418417
VERSION   AQ418417.1 GI:4476141
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
REFERENCE     Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominidae; Homo.
              1 (bases 1 to 328)
AUTHORS       Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
              Venter,J.C.
TITLE         Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
              Map Building
JOURNAL       Unpublished (1997)
COMMENT       Other GSSs: RPCI-11-177D15.TV
              Contact: Shaying Zhao, William Nierman, Mark Adams
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: hbe@tigr.org
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieter@dejong.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
              Research Genet cs (info@resgen.com). BAC end search page:
              http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
              Seq primer: SP6
              Class: BAC ends.
FEATURES      Location/Qualifiers
     source   1..328
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="GDB:7567670"
              /db_xref="taxon:9606"
              /clone="RPCI-11-177D15"
              /sex="Male"
              /cell_type="lymphocytes"
              /clone_lib="RPCI-11"
              /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
              RPCI11 Human Male BAC Library"
ORIGIN

Alignment Scores:
Pred. No.:              11.2          Length:          328
Score:                  11.00         Matches:          11
Percent Similarity:    100.0%         Conservative:      0
Best Local Similarity: 100.0%         Mismatches:        0
Query Match:             5.2%         Indels:            0
DB:                                   Gaps:              0

US-09-989-890-238 (1-212) x AQ418417 (1-328)

Qy   24 LeuProSerLeuAlaLeuLeuSerArgPro 34
         |||||||||||||||||||||||||||||
Db  259 CTTCCATCCAGCCTCGCTCTCCTCAGCCGTCCC 291

RESULT 34
BF115064/c
LOCUS       BF115064         257 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  hr73h01.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3134161 3'
            similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR. [1]
            ;, mRNA sequence.
ACCESSION   BF115064
VERSION     BF115064.1  GI:10984540
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 257)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 91.
FEATURES    Location/Qualifiers
     source  1..257
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:3134161"
            /lab_host="DH10B"
            /clone_lib="NCI_CGAP_Kid11"
            /note="Organ: kidney; Vector: p7T3D-Pac (Pharmacia) with
            a modified polylinker; Site_1: Not I; Site_2: Eco RI;
            Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
            prepared, and ss circles were made in vitro. Following HAP
            purification, this DNA was used as tracer in a subtractive
            hybridization reaction. The driver was PCR-amplified cDNAs
            from a pool of 5,000 clones made from the same library
            (cloneIDs 1322376-1323911, 1456007-1456775, and
            1500552-1502855). Subtraction by Bento Soares and M.
            Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:              72.4          Length:          257
Score:                  10.00         Matches:          10
Percent Similarity:    100.0%         Conservative:      0
Best Local Similarity: 100.0%         Mismatches:        0
Query Match:             4.7%         Indels:            0
DB:                                   Gaps:              0

US-09-989-890-238 (1-212) x BF115064 (1-257)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||
Db   92 GCGGGCCTGCCCCTCCGGCCTCGGCTGCCGCT 63

RESULT 35
AW197658/c
LOCUS       AW197658         277 bp    mRNA    linear   EST 29-NOV-1999
DEFINITION  xm85b08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2690967 3'
            similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR. [1]
            ;, mRNA sequence.
ACCESSION   AW197658
VERSION     AW197658.1  GI:6476888
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 277)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
```

```
                        Trace considered overall poor quality
                        Seq primer: -40UP from Gibco
                        High quality sequence stop: 1.
FEATURES                Location/Qualifiers
     source             1..277
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:2690967"
                        /lab_host="DH10B"
                        /clone_lib="NCI_CGAP_Kid11"
                        /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                        a modified polylinker; Site 1: Not I; Site 2: Eco RI;
                        Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                        prepared, and ss circles were made in vitro. Following HAP
                        purification, this DNA was used as tracer in a subtractive
                        hybridization reaction. The driver was PCR-amplified cDNAs
                        from a pool of 5,000 clones made from the same library
                        (cloneIDs 1322376-1323911, 1456007-1456775, and
                        1500552-1502855). Subtraction by Bento Soares and M.
                        Fatima Bonaldo. "

ORIGIN

Alignment Scores:
Pred. No.:              77.7              Length:        277
Score:                  10.00            Matches:       10
Percent Similarity:     100.0%           Conservative:  0
Best Local Similarity:  100.0%           Mismatches:    0
Query Match:            4.7%             Indels:        0
DB:                     1                Gaps:          0

US-09-989-890-238 (1-212) x AW197658 (1-277)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAlaAla 61
         |||||||||||||||||||||||||||||||||
Db  112 GCGGGCCTGCCCCTCGGCCTGGCTGCCGCT 83

RESULT 36
BF590030/c
LOCUS       BF590030           296 bp    mRNA      linear   EST 12-DEC-2000
DEFINITION  naa4h12.x1 NCI_CGAP_Kid1 Homo sapiens cDNA clone IMAGE:3259655 3'
            similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR.
            ;, mRNA sequence.
            1  (bases 1 to 296)
ACCESSION   BF590030
VERSION     BF590030.1  GI:11682354
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 296)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco.
FEATURES                Location/Qualifiers
     source             1..296
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
```

```
                        /clone="IMAGE:3259655"
                        /lab_host="DH10B"
                        /clone_lib="NCI_CGAP_Kid11"
                        /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                        a modified polylinker; Site 1: Not I; Site 2: Eco RI;
                        Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                        prepared, and ss circles were made in vitro. Following HAP
                        purification, this DNA was used as tracer in a subtractive
                        hybridization reaction. The driver was PCR-amplified cDNAs
                        from a pool of 5,000 clones made from the same library
                        (cloneIDs 1322376-1323911, 1456007-1456775, and
                        1500552-1502855). Subtraction by Bento Soares and M.
                        Fatima Bonaldo. "

ORIGIN

Alignment Scores:
Pred. No.:              82.7             Length:        296
Score:                  10.00           Matches:       10
Percent Similarity:     100.0%          Conservative:  0
Best Local Similarity:  100.0%          Mismatches:    0
Query Match:            4.7%            Indels:        0
DB:                     2               Gaps:          0

US-09-989-890-238 (1-212) x BF590030 (1-296)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAlaAla 61
         |||||||||||||||||||||||||||||||||
Db  130 GCGGGCCTGCCCCTCGGCCTGGCTGCCGCT 101

RESULT 37
AI382453/c
LOCUS       AI382453           297 bp    mRNA      linear   EST 18-MAR-1999
DEFINITION  ta72f03.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:2049629 3' similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN
            PROTEIN FKHR. ;, mRNA sequence.
ACCESSION   AI382453
VERSION     AI382453.1  GI:4195234
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 297)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 796     Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 248.
FEATURES                Location/Qualifiers
     source             1..297
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:2049629"
                        /dev_stage="8-9 weeks"
                        /lab_host="DH10B"
                        /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
                        was prepared from mRNA obtained from pooled 8-9 week
                        (total) fetus material with a Not I - oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGAGCGGCCGCTTAATTTTTTTTTTTTTTTTTT 3'].
                        Double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT7T3 vector. Library
                        went through one round of normalization, and was
```

```
ORIGIN              constructed by Bento Soares and M. Fatima Bonaldo. "

Alignment Scores:
Pred. No.:                83             Length:              297
Score:                    10.00          Matches:             10
Percent Similarity:       100.0%         Conservative:        0
Best Local Similarity:    100.0%         Mismatches:          0
Query Match:              4.7%           Indels:              0
DB:                                      Gaps:                0

US-09-989-890-238 (1-212) x AI382453 (1-297)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db    131 GCGGGCCTGCCCTCGGCCTCGGCTGCCGCT 102

RESULT 38
BY130650/c
LOCUS        BY130650                   312 bp    mRNA       linear   EST 08-DEC-2002
DEFINITION   BY130650 RIKEN full-length enriched, adult male brain Mus musculus
             cDNA clone L630092H16 5', mRNA sequence.
ACCESSION    BY130650
VERSION      BY130650.1  GI:26241751
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 312)
AUTHORS      Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
             Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
             Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
             Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
             Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
             Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
             Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
             Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
             Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
             Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
             Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
             Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
             Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
             Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
             Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
             Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
             Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
             Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
             Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
             Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
             Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
             Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
             Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
             Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
             Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
             Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
             Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE        Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
JOURNAL      Nature 420, 563-573 (2002)
PUBMED       12466851
COMMENT      Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center(GSC), Yokohama Institute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
             Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
             Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
             Miyazaki,A., Murata,M., Nakamura,K., Nomura,K., Numazaki,R.,
```

```
             Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
             Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
             Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
             Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new
             genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer. Genome Res.
             10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
             encyclopedia: real-time sequence clustering for construction of a
             nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
             further details.
FEATURES             Location/Qualifiers
     source          1..312
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="L630092H16"
                     /sex="male"
                     /tissue_type="brain"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male brain"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I"
ORIGIN

Alignment Scores:
Pred. No.:                87             Length:              312
Score:                    10.00          Matches:             10
Percent Similarity:       100.0%         Conservative:        0
Best Local Similarity:    100.0%         Mismatches:          0
Query Match:              4.7%           Indels:              0
DB:                                      Gaps:                0

US-09-989-890-238 (1-212) x BY130650 (1-312)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db     49 GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 20

RESULT 39
AL973065/c
LOCUS        AL973065                   322 bp    mRNA       linear   EST 27-NOV-2002
DEFINITION   AL973065 XGC-gastrula Xenopus tropicalis cDNA clone TGas12lp21 5',
             mRNA sequence.
ACCESSION    AL973065
VERSION      AL973065.1  GI:25796660
KEYWORDS     EST.
```

```
SOURCE      Xenopus tropicalis (western clawed frog)
ORGANISM    Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 322)
  AUTHORS   Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
  TITLE     Sanger Xenopus tropicalis EST project 2002
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Taylor R
            Sanger Institute
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS SEQUENCE ID: TGas12lp21.plkSP6
            Sequencing primer: SP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
FEATURES    Location/Qualifiers
     source 1..322
            /organism="Xenopus tropicalis"
            /mol_type="mRNA"
            /db_xref="taxon:8364"
            /clone="TGas12lp21"
            /dev_stage="gastrula (stages 10.5-12 mixed)"
            /lab_host="Escherichia coli XLi-blue"
            /clone_lib="XGC-gastrula"
            /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
            was oligo dT primed from 5ug of poly A+ RNA from stages
            10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
            into pCS107 with EcoRI at the 5' end and NotI at the 3'
            end."
ORIGIN

US-09-989-890-238 (1-212) x AL973065 (1-322)

Alignment Scores:
Pred. No.:                              Length:            322
Score:                 89.6             Matches:            10
                       10.00            Conservative:        0
Percent Similarity:    100.0%           Mismatches:          0
Best Local Similarity: 100.0%           Indels:              0
Query Match:           4.7%             Gaps:                0
DB:                    1

Qy   31  LeuSerArgProLeuSerProProProAla  40
         ||||||||||||||||||||||||||||||
Db  149  CTCTCCCGTCCCCTGTCCCCCCGCCCCTGCC  120

RESULT 40
BY125377/c
LOCUS       BY125377      325 bp    mRNA      linear   EST 08-DEC-2002
DEFINITION  BY125377 RIKEN full-length enriched, adult male brain Mus musculus
            cDNA clone L630036B03 5', mRNA sequence.
ACCESSION   BY125377
VERSION     BY125377.1  GI:26236478
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC) Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,N., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in Riken.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES    Location/Qualifiers
     source  1..325
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="L630036B03"
             /sex="male"
             /tissue_type="brain"
             /dev_stage="adult"
             /lab_host="DH10B"
             /clone_lib="RIKEN full-length enriched, adult male brain"
             /note="Site 1: SalI; Site 2: BamHI; cDNA library was
             prepared and sequenced in Mouse Genome Encyclopedia
             Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in
             RIKEN. Division of Experimental Animal Research in Riken
             contributed to prepare mouse tissues. 1st strand cDNA was
             primed with a primer [5'
             GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
             prepared by using trehalose thermo-activated reverse
```

LOCUS　　　　　BY336928　　　　　　　　　347 bp　　mRNA　　　linear　　EST 11-DEC-2002
DEFINITION　　BY336928 RIKEN full-length enriched, whole joints Mus musculus cDNA
　　　　　　　　clone L230004B21 5', mRNA sequence.
ACCESSION　　 BY336928
VERSION　　　　BY336928.1　GI:2653211
KEYWORDS　　　EST.
SOURCE　　　　 Mus musculus (house mouse)
ORGANISM　　　Mus musculus
　　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　　Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
　　　　　　　　Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE　　 1　(bases 1 to 347)
AUTHORS　　　 Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
　　　　　　　　Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
　　　　　　　　Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
　　　　　　　　Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
　　　　　　　　Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
　　　　　　　　Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
　　　　　　　　Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
　　　　　　　　Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
　　　　　　　　Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
　　　　　　　　Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
　　　　　　　　Kawaji,H., Kawasawa,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
　　　　　　　　Kurochkin,I.V., Lee,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
　　　　　　　　Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
　　　　　　　　Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
　　　　　　　　Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
　　　　　　　　Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
　　　　　　　　Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
　　　　　　　　Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
　　　　　　　　Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
　　　　　　　　Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
　　　　　　　　Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
　　　　　　　　Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
　　　　　　　　Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
　　　　　　　　Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
　　　　　　　　Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
　　　　　　　　Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
　　　　　　　　Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE　　　　 Analysis of the mouse transcriptome based on functional annotation
　　　　　　　　of 60,770 full-length cDNAs
JOURNAL　　　 Nature 420, 563-573 (2002)
PUBMED　　　　12466851
COMMENT　　　 Contact: Yoshihide Hayashizaki
　　　　　　　　Laboratory for Genome Exploration Research Group, RIKEN Genomic
　　　　　　　　Sciences Center(GSC), Yokohama Institute
　　　　　　　　The Institute of Physical and Chemical Research (RIKEN)
　　　　　　　　1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
　　　　　　　　Tel: 81-45-503-9222
　　　　　　　　Fax: 81-45-503-9216
　　　　　　　　Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
　　　　　　　　Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
　　　　　　　　Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
　　　　　　　　Miyazaki,A., Murata,M., Nakamura,Y., Nomura,K., Numazaki,R.,
　　　　　　　　Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
　　　　　　　　Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
　　　　　　　　Hayashizaki,Y. Direct Submission
　　　　　　　　Computational Analysis of Full-Length Mouse cDNAs Compared with
　　　　　　　　Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
　　　　　　　　Normalization and subtraction of cap-trapper-selected cDNAs to
　　　　　　　　prepare full-length cDNA libraries for rapid discovery of new
　　　　　　　　genes. Genome Res. 10 (10), 1617-1630 (2000)
　　　　　　　　RIKEN integrated sequence analysis (RISA) system--384-format
　　　　　　　　sequencing pipeline with 384 multicapillary sequencer. Genome Res.
　　　　　　　　10 (11), 1757-1771 (2000)
　　　　　　　　Computer-based methods for the mouse full-length cDNA
　　　　　　　　encyclopedia: real-time sequence clustering for construction of a
　　　　　　　　nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
　　　　　　　　cDNA library was prepared and sequenced in Mouse Genome
　　　　　　　　Encyclopedia Project of Genome Exploration Research Group in Riken
　　　　　　　　Genomic Sciences Center and Genome Science Laboratory in RIKEN.
　　　　　　　　Division of Experimental Animal Research in Riken contributed to
　　　　　　　　prepare mouse tissues.

ORIGIN
　　　　　　transcriptase and subsequently enriched for full-length by
　　　　　　cap-trapper. Second strand cDNA was prepared with the
　　　　　　primer adapter of sequence [5'
　　　　　　GAGAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA
　　　　　　was cloned into the XhoI and BamHI sites. Vector: a
　　　　　　modified pBluescript KS(+) after bulk excision from Lambda
　　　　　　FLC I"

Alignment Scores:
Pred. No.:　　　90.4　　　　　Length:　　　　　　325
Score:　　　　　10.00%　　　　Matches:　　　　　 10
Percent Similarity:　100.0%　　Conservative:　　　0
Best Local Similarity:　100.0%　Mismatches:　　　　0
Query Match:　　　4.7%　　　　Indels:　　　　　　 0
DB:　　　　　　　5　　　　　　Gaps:　　　　　　 0

US-09-989-890-238 (1-212) x BY125377 (1-325)

Qy　　52 AlaGlyLeuProSerAlaSerAlaAlaAlaAla 61
　　　　 ||||||||||||||||||||||||||||||||||
Db　　49 GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 20

RESULT 41
BP423898
LOCUS　　　　 BP423898　　　　　　341 bp　　mRNA　　　linear　　EST 27-MAY-2005
DEFINITION　 BP423898 Homo sapiens small intestine Homo sapiens cDNA clone
　　　　　　　 HIB07478r 3', mRNA sequence.
ACCESSION　　BP423898
VERSION　　　 BP423898.1　GI:66790151
KEYWORDS　　　EST.
SOURCE　　　　Homo sapiens (human)
ORGANISM　　　Homo sapiens
　　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　　Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
　　　　　　　　Hominidae; Homo.
REFERENCE　　 1　(bases 1 to 341)
AUTHORS　　　 Takeda,J., Jin,L. and Horikawa,Y.
TITLE　　　　 Expression profile of mRNAs from human pancreatic islet tumors
JOURNAL　　　 Unpublished (2005)
COMMENT　　　 Contact: Yukio Horikawa
　　　　　　　　Laboratory of Molecular Genetics
　　　　　　　　Institute for Molecular and Cellular Regulation, Gunma University
　　　　　　　　Showa-machi 3-39-15, Maebashi, Gunma, 371-8512, Japan
　　　　　　　　Tel: 81-27-220-8832
　　　　　　　　Fax: 81-27-220-8889
　　　　　　　　Email: yhorikaw@showa.gunma-u.ac.jp.
FEATURES　　　 Location/Qualifiers
　　source　　　1..341
　　　　　　　　/organism="Homo sapiens"
　　　　　　　　/mol_type="mRNA"
　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　/clone="HIB07478r"
　　　　　　　　/tissue_type="small intestine"
　　　　　　　　/clone_lib="Homo sapiens small intestine"
ORIGIN

Alignment Scores:
Pred. No.:　　　94.6　　　　　Length:　　　　　　341
Score:　　　　　10.00　　　　Matches:　　　　　 10
Percent Similarity:　100.0%　　Conservative:　　　0
Best Local Similarity:　100.0%　Mismatches:　　　　0
Query Match:　　　4.7%　　　　Indels:　　　　　　 0
DB:　　　　　　　3　　　　　　Gaps:　　　　　　 0

US-09-989-890-238 (1-212) x BP423898 (1-341)

Qy　　52 AlaGlyLeuProSerAlaSerAlaAlaAlaAla 61
　　　　 ||||||||||||||||||||||||||||||||||
Db　　210 GCGGGCCTGCCCTCCGGCCCTCGGCTGCCGCT 239

RESULT 42
BY336928/c

```
JOURNAL     Nature 420, 563-573 (2002)
PUBMED      12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC) Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in RIKEN.
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Vassilis Aidinis ( Biomedical Sciences
            Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
            Fleming street 16672 Vari,Greece ) whose assistance we gratefully
            acknowledge.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..349
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="L230015C04"
                     /tissue_type="whole joints"
                     /clone_lib="RIKEN full-length enriched, whole joints"
ORIGIN

Alignment Scores:
Pred. No.:             1..349
Score:                 96.6                 Length:          349
Percent Similarity:    10.00                Matches:         10
Best Local Similarity: 100.0%               Conservative:    0
Query Match:           100.0%               Mismatches:      0
DB:                    4.7%                 Indels:          0
                       5                    Gaps:            0

US-09-989-890-238 (1-212) x BY338512 (1-349)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db     73 GCTGGGCTGCCTAGTGCCAGCGCGTGCTGCG 44
```

RESULT 44
```
BY311584/c
LOCUS       BY311584                 350 bp    mRNA    linear   EST 11-DEC-2002
DEFINITION  BY311584 RIKEN full-length enriched, stroma cell Mus musculus cDNA
            clone I320021G07 5', mRNA sequence.
ACCESSION   BY311584
VERSION     BY311584.1  GI:26501921
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
                     Tissues were provided by Vassilis Aidinis ( Biomedical Sciences
            Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
            Fleming street 16672 Vari,Greece ) whose assistance we gratefully
            acknowledge.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..347
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="L230004B21"
                     /tissue_type="whole joints"
                     /clone_lib="RIKEN full-length enriched, whole joints"
ORIGIN

Alignment Scores:
Pred. No.:             96.1                 Length:          347
Score:                 10.00                Matches:         10
Percent Similarity:    100.0%               Conservative:    0
Best Local Similarity: 100.0%               Mismatches:      0
Query Match:           4.7%                 Indels:          0
DB:                    5                    Gaps:            0

US-09-989-890-238 (1-212) x BY336928 (1-347)

Qy     52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db     73 GCTGGGCTGCCTAGTGCCAGCGCGTGCTGCG 44
```

RESULT 43
```
BY338512/c
LOCUS       BY338512                 349 bp    mRNA    linear   EST 11-DEC-2002
DEFINITION  BY338512 RIKEN full-length enriched, whole joints Mus musculus cDNA
            clone L230015C04 5', mRNA sequence.
ACCESSION   BY338512
VERSION     BY338512.1  GI:26534978
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogani,A.,
            Schonbach,C., Gojobori,T., Baidarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
```

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 350)
AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.B., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
PUBMED   12466651
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K. Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Takashi Ishikawa (Department of Surgery
            2 Yokohama City University 3-9 Fukuura,Kanazawa-ku,Yokohama
            236-0004 Japan) whose assistance we gratefully acknowledge.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..350
                     /organism="Mus musculus"
                     /mol_type="mRNA"

                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="I320021G07"
                     /cell_type="stroma cell"
                     /clone_lib="RIKEN full-length enriched, stroma cell"
ORIGIN
Alignment Scores:
Pred. No.:   96.9   Length: 350
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
DB: 4.7%   Indels: 0
                                       Gaps: 0

US-09-989-890-238 (1-212) x BY311584 (1-350)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAlaAla 61
         ||||||||||||||||||||||||||||||||
Db   49 GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 20

RESULT 45
CD598944
LOCUS   CD598944   350 bp   mRNA   linear   EST 16-JUN-2003
DEFINITION RK137A3B12.T3 Zebrafish Kidney Marrow cDNA library Danio rerio cDNA
            clone RK137A3B12 5', mRNA sequence.
ACCESSION   CD598944
VERSION   CD598944.1 GI:31783010
KEYWORDS   EST.
SOURCE   Danio rerio (zebrafish)
  ORGANISM Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 350)
AUTHORS   Song,H.D., Sun,X.J., Deng,M., Zhang,G.W., Zhou,Y., Wu,X.Y.,
            Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y., Zon,L.I.,
            Kanki,J.P., Liu,T.X., Look,A.T. and Chen,Z.
TITLE   Hematopoietic gene expression profile in zebrafish kidney marrow
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (46), 16240-16245 (2004)
PUBMED   15520368
COMMENT   Contact: Chen Z.
            State Key Lab for Medical Genomics
            Shanghai Institute of Hematology, Ruijin Hospital Affiliated to
            Shanghai Second Medical University
            197 Rui Jin Road II, Shanghai 200025, P. R. China
            Tel: 86-21-64740490
            Fax: 86-21-6474320g
            Email: zchen@stn.sh.cn
            Seq primer: T3.
FEATURES   Location/Qualifiers
     source   1..350
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="RK137A3B12"
                     /dev_stage="mature"
                     /clone_lib="Zebrafish Kidney Marrow cDNA library"
                     /note="Organ: kidney; Vector: pBS-CMV; Site_1: XhoI;
                     Site_2: EcoRI; Total RNA was extracted from the kidney
                     tissues of mature zebrafish. The poly (A)+ RNA fraction
                     was separated from total RNA by oligo (dT) cellulose
                     chromatography. Library was initially constructed in the
                     lambdaZAP Express vector (Stratagene) and in vivo excised
                     into pBS-CMV vector."
ORIGIN
Alignment Scores:
Pred. No.:   96.9   Length: 350
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0

```
DB:                              6        Gaps:             0

US-09-989-890-238 (1-212) x CD598944 (1-350)

Qy      27  SerLeuAlaLeuLeuSerArgProLeuSer 36
            ||||||||||||||||||||||||||||||
Db     254  AGCCTAGCATTGTTATCAAGACCGTTGAGC 283

RESULT 46
BY343852/c
LOCUS           BY343852                352 bp    mRNA        linear   EST 12-DEC-2002
DEFINITION      BY343852 RIKEN full-length enriched, whole joints Mus musculus cDNA
                clone L230047L01 5', mRNA sequence.
ACCESSION       BY343852
VERSION         BY343852.1  GI:26573340
KEYWORDS        EST.
SOURCE          Mus musculus (house mouse)
ORGANISM        Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE       1  (bases 1 to 352)
AUTHORS         Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
                Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
                Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
                Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
                Hume,D.A., Quackenbush,J., Schrim,L.M., Kanapin,A., Matsuda,H.,
                Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
                Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
                Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
                Gariboldi,M., Gissi,C., Godzik,C., Gough,J., Grimmond,S.,
                Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
                Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
                Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
                Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
                Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
                Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
                Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
                Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
                Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
                Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
                Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
                Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
                Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
                Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
                Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
                Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
                Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
                Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE           Analysis of the mouse transcriptome based on functional annotation
                of 60,770 full-length cDNAs
JOURNAL         Nature 420, 563-573 (2002)
PUBMED          12466851
COMMENT         Contact: Yoshihide Hayashizaki
                Laboratory for Genome Exploration Research Group, RIKEN Genomic
                Sciences Center(GSC), Yokohama Institute
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81-45-503-9222
                Fax: 81-45-503-9216
                Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
                Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
                Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
                Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
                Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
                Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
                Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
                Hayashizaki,Y. Direct Submission
                Computational Analysis of Full-Length Mouse cDNAs Compared with
                Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
                Normalization and subtraction of cap-trapper-selected cDNAs to
                prepare full-length cDNA libraries for rapid discovery of new
                genes. Genome Res. 10 (10), 1617-1630 (2000)
                RIKEN integrated sequence analysis (RISA) system-384-format
```

```
                sequencing pipeline with 384 multicapillary sequencer. Genome Res.
                10 (11), 1757-1771 (2000)
                Computer-based methods for the mouse full-length cDNA
                encyclopedia: real-time sequence clustering for construction of a
                nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
                cDNA library was prepared and sequenced in Mouse Genome
                Encyclopedia Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in RIKEN.
                Division of Experimental Animal Research in Riken contributed to
                prepare mouse tissues.
                Tissues were provided by Vassilis Aidinis ( Biomedical Sciences
                Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
                Fleming street 16672 Vari,Greece ) whose assistance we gratefully
                acknowledge.
                Please visit our web site (http://genome.gsc.riken.go.jp) for
                further details.
FEATURES             Location/Qualifiers
     source          1..352
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="L230047L01"
                     /tissue_type="whole joints"
                     /clone_lib="RIKEN full-length enriched, whole joints"
ORIGIN

Alignment Scores:
Pred. No.:              97.4                      Length:             352
Score:                  10.00                     Matches:             10
Percent Similarity:     10.00                     Conservative:         0
Best Local Similarity:  100.0%                    Mismatches:           0
Query Match:            4.7%                       Indels:               0
DB:                     5                         Gaps:                 0

US-09-989-890-238 (1-212) x BY343852 (1-352)

Qy      52  AlaGlyLeuProSerAlaSerArgAlaAlaAla 61
            |||||||||||||||||||||||||||||||||
Db      73  GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 44

RESULT 47
BY338857/c
LOCUS           BY338857                355 bp    mRNA        linear   EST 11-DEC-2002
DEFINITION      BY338857 RIKEN full-length enriched, whole joints Mus musculus cDNA
                clone L230018D18 5', mRNA sequence.
ACCESSION       BY338857
VERSION         BY338857.1  GI:26535539
KEYWORDS        EST.
SOURCE          Mus musculus (house mouse)
ORGANISM        Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE       1  (bases 1 to 355)
AUTHORS         Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
                Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
                Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
                Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
                Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
                Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
                Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
                Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
                Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
                Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
                Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
                Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
                Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
                Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
                Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
                Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
                Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
                Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
                Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
```

Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,A., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,B.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
PUBMED   12466851
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Vassilis Aidinis (Biomedical Sciences
            Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
            Fleming street 16672 Vari,Greece) whose assistance we gratefully
            acknowledge.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..355
                 /organism="Mus musculus"
                 /mol_type="mRNA"
                 /db_xref="taxon:10090"
                 /clone="L230018D18"
                 /tissue_type="whole joints"
                 /clone_lib="RIKEN full-length enriched, whole joints"
ORIGIN

Alignment Scores:
Pred. No.:   98.2   Length: 355
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x BY338857 (1-355)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   73 GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 44

RESULT 48

BF733006/c
LOCUS   BF733006   362 bp   mRNA   linear   EST 30-MAR-2001
DEFINITION nae07d09.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3434560 3'
            similar to SW:FKHR_HUMAN Q12778 FORK-HEAD DOMAIN PROTEIN FKHR. [1]
            ;, mRNA sequence.
ACCESSION   BF733006
VERSION   BF733006.1 GI:1205081
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 362)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 349.
FEATURES   Location/Qualifiers
     source   1..362
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:3434560"
                 /tissue_type="fibrotheoma"
                 /lab_host="DH10B (phage-resistant)"
                 /clone_lib="NCI_CGAP_Ov18"
                 /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
                 strand cDNA was primed with a Not I - oligo(dT) primer [5'
                 TGTTACCAATCTGAAGTGGACGGCGCGGCGACCATTTTTTTTTTTTTTTTTT 3'];
                 double-stranded cDNA was ligated to Eco RI adaptors
                 (Pharmacia), digested with Not I and cloned into the Not
                 I and Eco RI sites of the modified pT7T3 vector. Library
                 went through one round of normalization, and was
                 constructed by Bento Soares and M. Fatima Bonaldo.   "
ORIGIN

Alignment Scores:
Pred. No.:   Length: 362
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x BF733006 (1-362)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   131 GCGGGCCTGCCCTGCCTCGGCGTGCCGCT 102

RESULT 49
AI631631/c
LOCUS   AI631631   365 bp   mRNA   linear   EST 16-DEC-1999
DEFINITION wb01c08.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2304398 3'
            similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR. [1]
            ;, mRNA sequence.
ACCESSION   AI631631
VERSION   AI631631.1 GI:482961
KEYWORDS   EST.

```
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 365)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: M. Bento Soares, Ph.D., M. Fatima
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 767   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 324.
FEATURES             Location/Qualifiers
     source          1..365
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2304398"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GC6"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

US-09-989-890-238 (1-212) x AI631631 (1-365)

Alignment Scores:
Pred. No.:           101            Length:         365
Score:               10.00          Matches:        10
Percent Similarity:  100.0%         Conservative:   0
Best Local Similarity: 100.0%       Mismatches:     0
Query Match:         4.7%           Indels:         0
DB:                  1              Gaps:           0

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAlaAla 61
            |||||||||||||||||||||||||||||||
Db    99 GCGGGCCTGCCCTGCCCTCGGCTGCCGCT 70

RESULT 50
BY019540/c
LOCUS       BY019540             388 bp    mRNA      linear     EST 06-DEC-2002
DEFINITION  BY019540 RIKEN full-length enriched, mammary gland RCB-0526
            Jyg-MC(A) cDNA Mus musculus cDNA clone G830035L04 5', mRNA
            sequence.
ACCESSION   BY019540
VERSION     BY019540.1  GI:26079789
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 388)
```

```
AUTHORS     Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
PUBMED      12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..388
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="G830035L04"
                     /tissue_type="mammary gland"
                     /cell_line="RCB-0526 Jyg-MC(A)"
                     /clone_lib="RIKEN full-length enriched, mammary gland
                     RCB-0526 Jyg-MC(A) cDNA"
```

RESULT 51

ORIGIN

Alignment Scores:
Pred. No.: 107
Score: 10.00   Length: 388
Percent Similarity: 100.0%   Matches: 10
Best Local Similarity: 100.0%   Conservative: 0
Query Match: 4.7%   Mismatches: 0
DB: 5   Indels: 0
                                  Gaps: 0

US-09-989-890-238 (1-212) x BY019540 (1-388)

```
Qy  52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
       ||||||||||||||||||||||||||||||
Db  76 GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 47
```

RESULT 51
BB863674/c
LOCUS   BB863674   401 bp   mRNA   linear   EST 29-SEP-2000
DEFINITION UI-M-BH0-akc-f-05-0-UI.r1 NIH_BMAP_M_S1 Mus musculus cDNA clone
            UI-M-BH0-akc-f-05-0-UI 5', mRNA sequence.
ACCESSION   BB863674
VERSION   BB863674.1 GI:10383952
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 401)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
PUBMED   8889548
COMMENT   Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1..401
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UI-M-BH0-akc-f-05-0-UI"
                    /dev_stage="27-32 days"
                    /lab_host="DH10B (Life Technologies)"
                    /clone_lib="NIH_BMAP_M_S1"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; The
                    NIH_BMAP_M_S1 library is a subtracted library derived from
                    a mixture of normalized libraries from ten regions of the
                    mouse brain (cerebellum, brain stems, olfactory bulbs,
                    hypothalamus, cortex, amygdala, basal ganglia, pineal
                    gland, striatum, hippocampus). The driver used for
                    subtraction consisted of a pool of 20,000 cDNA clones
                    obtained from non-normalized and normalized libraries of
                    these ten regions of the mouse brain."
ORIGIN

Alignment Scores:
Pred. No.: 110   Length: 401
Score:

RESULT 52

ORIGIN

Alignment Scores:
Pred. No.: 111
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 2
DB: Gaps: 0

US-09-989-890-238 (1-212) x BB863674 (1-401)

```
Qy  52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
       ||||||||||||||||||||||||||||||
Db 117 GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 88
```

RESULT 52
CA949507/c
LOCUS   CA949507   404 bp   mRNA   linear   EST 31-DEC-2002
DEFINITION iq25c06.x1 HR85 islet Homo sapiens cDNA clone IMAGE: 3' similar to
            SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR. [1] ; mRNA
            sequence.
ACCESSION   CA949507
VERSION   CA949507.1 GI:27442384
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 404)
AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmit,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
TITLE   Endocrine Pancreas Consortium
JOURNAL   Unpublished (2000)
COMMENT   Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 401.
FEATURES   Location/Qualifiers
     source   1..404
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:"
                    /tissue_type="Purified pancreatic islet"
                    /lab_host="DH10B"
                    /clone_lib="HR85 islet"
                    /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                    NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                    Size-selected on agarose gel. Average insert size -1kb. 5'
                    XhoI site was destroyed after directional cloning.
                    Amplified once. Contact information: Hiroshi Inoue, MD,
                    Metabolism Div. (Alan Permutt Lab). Washington University
                    School of Medicine, Box 8127, 660 South Euclid Ave., St.
                    Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                    314-362-1916, Fax: 314-747-2692."
ORIGIN

Alignment Scores:
Pred. No.:   Length: 404
Score: 111   Matches: 10
Percent Similarity: 100.0%   Conservative: 0

```
Best Local Similarity:    100.0%          Mismatches:    0
Query Match:                4.7%          Indels:        0
DB:                           6           Gaps:          0

US-09-989-890-238 (1-212) x CA949507 (1-404)

Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||||
Db      97 GCGGGCCTGCCCTCGGCCTCGGCTGCCGCT 68
```

```
RESULT 53
AI494178/c
LOCUS        AI494178       419 bp    mRNA    linear    EST 13-APR-1999
DEFINITION   ti14g02.y1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2130482 5'
             similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR. ;,
             mRNA sequence.
ACCESSION    AI494178
VERSION      AI494178.1   GI:4395181
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE    1  (bases 1 to 419)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
             Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 790   Std Error: 0.00
             Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1..419
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2130482"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Kid11"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site 1: Not I; Site 2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and as circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
ORIGIN
Alignment Scores:
Pred. No.:                  115          Length:       419
Score:                    10.00          Matches:       10
Percent Similarity:       100.0%         Conservative:   0
Best Local Similarity:    100.0%         Mismatches:     0
Query Match:                4.7%         Indels:         0
DB:                           1          Gaps:           0

US-09-989-890-238 (1-212) x AI494178 (1-419)

Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||||
```

```
Db     122 GCGGGCCTGCCCTCGGCCTCGGCTGCCGCT 93

RESULT 54
AI854714
LOCUS        AI854714       421 bp    mRNA    linear    EST 15-JUL-1999
DEFINITION   UI-M-BH0-akc-f-05-0-UI.s1 NIH_BMAP_M_S1 Mus musculus cDNA clone
             UI-M-BH0-akc-f-05-0-UI 3', mRNA sequence.
ACCESSION    AI854714
VERSION      AI854714.1   GI:5498620
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 421)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
             discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
PUBMED       8889548
COMMENT      Contact: Chin, H
             National Institute of Mental Health
             6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
             20892-9643, USA
             Tel: 301 443 1706
             Fax: 301 443 9890
             Email: mEST@mail.nih.gov
             The sequence contained an oligo-dT track that was present in the
             oligonucleotide that was used to prime the synthesis of first
             strand cDNA and therefore this may represent a bonafide poly A
             tail. The sequence tag present in the cDNA between the NotI site
             and the oligo-dT track served to identify it as a clone from the
             normalized corpus striatum cDNA library Preparation: M.B.
             Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
             available by the means that is soon to be determined. When NIH
             determines the means for distribution of the BMAP cDNA clones, this
             record will be updated accordingly when that means is determined.
             Seq primer: M13 Forward
             POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..421
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH0-akc-f-05-0-UI"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NIH_BMAP_M_S1"
                     /note="vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; The
                     NIH_BMAP_M_S1 library is a subtracted library derived from
                     a mixture of normalized libraries from ten regions of the
                     mouse brain (cerebellum, brain stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal ganglia, pineal
                     gland, striatum, hippocampus). The driver used for
                     subtraction consisted of a pool of 20,000 cDNA clones
                     obtained from non-normalized and normalized libraries of
                     these ten regions of the mouse brain.
                     TAG_TISSUE=corpus-striatum
                     TAG_LIB=NIH_BMAP_M_S1
                     TAG_SEQ=ACGGC"
ORIGIN
Alignment Scores:
Pred. No.:                  115          Length:       421
Score:                    10.00          Matches:       10
Percent Similarity:       100.0%         Conservative:   0
Best Local Similarity:    100.0%         Mismatches:     0
Query Match:                4.7%         Indels:         1
DB:                                      Gaps:           0
```

US-09-989-890-238 (1-212) x AI854714 (1-421)

```
Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         |||||||||||||||||||||||||||||
Db   323 GCTGGGCTGCCCTAGTGCCAGCGCGCTGCGC 352
```

RESULT 55
AI188507/c
LOCUS   AI188507   434 bp   mRNA   linear   EST 28-OCT-1998
DEFINITION qd14e10.x1 Soares placenta 8to9weeks 2NbHP8to9W Homo sapiens cDNA
            clone IMAGE:1723722 3' similar to SW:FKHR_HUMAN Q12778 FORK HEAD
            DOMAIN PROTEIN FKHR. ; mRNA sequence.
ACCESSION   AI188507
VERSION   AI188507.1 GI:3739716
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 434)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 713   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 417.
FEATURES   Location/Qualifiers
source   1..434
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:1723722"
            /dev_stage="two placentae: one from 8 weeks and another
            from 9 weeks post conception"
            /lab_host="DH10B (ampicillin resistant)"
            /clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
            /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
            modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGGCGGCCGCGAGTTTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was size selected, ligated to Eco RI
            adapters (Pharmacia), digested with Not I and cloned into
            the Not I and Eco RI sites of a modified pT7T3 vector
            (Pharmacia). Library constructed by Bento Soares and
            M.Fatima Bonaldo."

ORIGIN

Alignment Scores:
Pred. No.:   119   Length: 434
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 1   Gaps: 0

US-09-989-890-238 (1-212) x AI188507 (1-434)

```
Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         |||||||||||||||||||||||||||||
Db   131 GCGGGGCCTGCCCTCGGCCTCGGCGTGCCGCT 102
```

RESULT 56
AW467069/c
LOCUS   AW467069   459 bp   mRNA   linear   EST 24-FEB-2000
DEFINITION ha09a02.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2873162 3'
            similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR. [1]
            ; mRNA sequence.

ACCESSION   AW467069
VERSION   AW467069.1 GI:7037175
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 459)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 445.
FEATURES   Location/Qualifiers
source   1..459
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2873162"
            /tissue_type="2 pooled tumors (clear cell type)"
            /lab_host="DH10B"
            /clone_lib="NCI_CGAP_Kid12"
            /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
            a modified polylinker; Site 1: Not I; Site 2: Eco RI;
            Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
            prepared, and ss circles were made in vitro. Following HAP
            purification, this DNA was used as tracer in a subtractive
            hybridization reaction. The driver was PCR-amplified cDNAs
            from a pool of 5,000 clones made from the same library
            (cloneIDs 1323912-1325831, 1471368-1472903 and
            1492104-1493255). Subtraction by Bento Soares and M.
            Fatima Bonaldo. "

ORIGIN

Alignment Scores:
Pred. No.:   125   Length: 459
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 1   Gaps: 0

US-09-989-890-238 (1-212) x AW467069 (1-459)

```
Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         |||||||||||||||||||||||||||||
Db    90 GCGGGGCCTGCCCTCGGCCTCGGCGTGCCGCT 61
```

RESULT 57
AW420053
LOCUS   AW420053   460 bp   mRNA   linear   EST 09-FEB-2000
DEFINITION fj86g08.y1 zebrafish gridded kidney Danio rerio cDNA 5', mRNA
            sequence.
ACCESSION   AW420053
VERSION   AW420053.1 GI:6947985
KEYWORDS   EST.
SOURCE   Danio rerio (zebrafish)
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS   Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
            1 (bases 1 to 460)
            Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
            Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
TITLE   WashU Zebrafish EST Project 1998
JOURNAL   Unpublished (1998)
COMMENT   Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            Genome Systems, St. Louis, Missouri (web address:
            www.genomesystems.com) (email contact: info@genomesystems.com) and
            Research Genetics, Huntsville, Alabama (web address:
            www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 426.
FEATURES   Location/Qualifiers
     source   1..460
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /sex="mixed"
                     /tissue_type="kidney pooled from 300 wild type adults"
                     /lab_host="XLOLR"
                     /clone_lib="zebrafish gridded kidney"
                     /note="Organ: kidney; Vector: pBK-CMV; Site_1: EcoRI;
                     Site_2: XhoI; Oligo dT cDNA library constructed from mRNA
                     pooled from pooled kidney tissue from 300 adult
                     zebrafish."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 460
Score: 125.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x AW420053 (1-460)

Qy   27 SerLeuAlaLeuLeuSerArgProLeuSer 36
          ||||||||||||||||||||||||||||||
Db   223 AGCCTAGCATTGTTATCAAGACCGTTGAGC 252

RESULT 58
CA941290/c
LOCUS   CA941290   488 bp   mRNA   linear   EST 30-DEC-2002
DEFINITION ir33b08.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6546952 3',
            mRNA sequence.
            similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR. [1]
            ;, mRNA sequence.
ACCESSION   CA941290
VERSION   CA941290.1 GI:27429770
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 488)
AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,

---

            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
TITLE   Endocrine Pancreas Consortium
JOURNAL   Unpublished (2000)
COMMENT   Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 439.
FEATURES   Location/Qualifiers
     source   1..488
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6546952"
                     /tissue_type="Purified pancreatic islet"
                     /clone_lib="HR85 islet"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, B-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 488
Score: 133.00   Matches: 10
Percent Similarity: 100.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x CA941290 (1-488)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
          ||||||||||||||||||||||||||||||
Db   86 GCGGGCCTGCCCTCGGCCTCCGGCTGCCGCT 57

RESULT 59
AV780819/c
LOCUS   AV780819   504 bp   mRNA   linear   EST 19-AUG-2004
DEFINITION AV780819 Lotus japonicus Pods (20-30 mm in length) Lotus
            corniculatus var. japonicus cDNA clone MPDL083d04_f 3', mRNA
            sequence.
ACCESSION   AV780819
VERSION   AV780819.1 GI:45404904
KEYWORDS   EST.
SOURCE   Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1 (bases 1 to 504)
AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE   Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis

```
JOURNAL      Plant Mol. Biol. 54 (3), 405-414 (2004)
PUBMED       1528495
COMMENT      Contact: Erika Asamizu
             The First Laboratory for Plant Gene Research
             Kazusa DNA Research Institute
             Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
             Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES     Location/Qualifiers
   source    1..504
             /organism="Lotus corniculatus var. japonicus"
             /mol_type="mRNA"
             /isolate="Miyakojima MG-20"
             /db_xref="taxon:34305"
             /clone="MPDL083d04_f"
             /tissue_type="Pods (20-30 mm in length)"
             /clone_lib="Lotus japonicus Pods (20-30 mm in length)"
             /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
             XhoI"
ORIGIN

Alignment Scores:
Pred. No.:                 137          Length:              504
Score:                     10.00        Matches:             10
Percent Similarity:        100.0%       Conservative:        0
Best Local Similarity:     100.0%       Mismatches:          0
Query Match:               4.7%         Indels:              0
DB:                        1            Gaps:                0

US-09-989-890-238 (1-212) x AV780819 (1-504)

Qy      23 AlaLeuProSerSerLeuAlaLeuLeuSer 32
           ||||||||||||||||||||||||||||||
Db     215 GCACTTCCATCTTCCCTTGCCCTGCTCTCA 186

RESULT 60
BF434989/c
LOCUS        BF434989         511 bp    mRNA    linear    EST 29-NOV-2000
DEFINITION   7p04c04.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644910 3'
             similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR. [1]
             ;contains MER22.t2 TAR1 repetitive element ;, mRNA sequence.
ACCESSION    BF434989
VERSION      BF434989.1  GI:11447277
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
             Hominidae; Homo.
REFERENCE    1  (bases 1 to 511)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
             R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
             Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
             I.M.A.G.E. Consortium DNA Sequencing by: Washington University
             Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL, send email to:
             info@image.llnl.gov
             Seq primer: -40UP from Gibco
             High quality sequence stop: 459.
FEATURES     Location/Qualifiers
   source    1..511
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:3644910"
             /tissue_type="fibrotheoma"
             /lab_host="DH10B (phage-resistant)"
```

```
             /clone_lib="NCI_CGAP_Ov18"
             /note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a
             modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
             strand cDNA was primed with a Not I - oligo(dT) primer [5'
             TGTTACCAATCTGAAGTGGAGCGCCCGGGACAATTTTTTTTTTTTTTTT 3'];
             double-stranded cDNA was ligated to Eco RI adaptors
             (Pharmacia), digested with Not I and cloned into the Not
             I and Eco RI sites of the modified pT7T3 vector. Library
             went through one round of normalization, and was
             constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:                 138          Length:              511
Score:                     10.00        Matches:             10
Percent Similarity:        100.0%       Conservative:        0
Best Local Similarity:     100.0%       Mismatches:          0
Query Match:               4.7%         Indels:              0
DB:                        2            Gaps:                0

US-09-989-890-238 (1-212) x BF434989 (1-511)

Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla Ala 61
           ||||||||||||||||||||||||||||||||||
Db      85 GCGGGCCTGCCCTCGGCCTCGGCTGCCGCT 56

RESULT 61
BH298354
LOCUS        BH298354                 557 bp    DNA    linear    GSS 30-NOV-2001
DEFINITION   CH230-146C24.TVB CHORI-230 Segment 1 Rattus norvegicus genomic
             clone CH230-146C24, genomic survey sequence.
ACCESSION    BH298354
VERSION      BH298354.1  GI:17210762
KEYWORDS     GSS.
SOURCE       Rattus norvegicus (Norway rat)
ORGANISM     Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE    1  (bases 1 to 557)
AUTHORS      Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
             Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
             Riggs,F., de Jong,P. and Fraser,C.M.
TITLE        Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL      Unpublished (1999)
COMMENT      Contact: Shaying Zhao
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0200
             Fax: 301 838 0208
             Email: szhao@tigr.org
             Clones are derived from the rat BAC library CHORI-230
             (http://www.chori.org/bacpac/rat230.htm). For BAC library
             availability, please contact Pieter de Jong (pdejong@mail.cho.org).
             Clones may be purchased from BACPAC Resources
             (http://www.chori.org/bacpac/or ering_information.htm). BAC end
             page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
             Plate: 146  row: C  column: 24
             Seq primer: T7
             Class: BAC ends.
FEATURES     Location/Qualifiers
   source    1..557
             /organism="Rattus norvegicus"
             /mol_type="genomic DNA"
             /strain="BN/SsNHsd/MCW"
             /db_xref="taxon:10116"
             /clone="CH230-146C24"
             /sex="Female"
             /cell_type="Brain"
             /clone_lib="CHORI-230 Segment 1"
             /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
             CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
```

ORIGIN   Pieter de Jong"

Alignment Scores:
Pred. No.:   150   Length: 557
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x BH298354 (1-557)

Qy   25 ProSerSerLeuAlaLeuLeuSerArgPro 34
             |||||||||||||||||||||||||||||||
Db   204 CCATCCAGCTTAGCTCTTCTCTCCAGACCC 233

RESULT 62
CC740775/c
LOCUS   CC740775   560 bp   DNA   linear   GSS 25-JUN-2003
DEFINITION   ZMMBBb0108F12.r ZMMBBb Zea mays genomic clone ZMMBBb0108F12 3',
              genomic survey sequence.
ACCESSION   CC740775
VERSION   CC740775.1 GI:32193228
KEYWORDS   GSS.
SOURCE   Zea mays
ORGANISM   Zea mays
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
              clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 560)
AUTHORS   Yu,Y., Kim,H.R., Hatfield,J., Soderlund,C., Bharti,A.K., Messing,J.
              and Wing,R.
TITLE   Sequencing of the maize genome
JOURNAL   Unpublished (2003)
COMMENT   Contact: Rod Wing
              Arizona Genomics Institute
              University of Arizona
              Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
              85721-0088, USA
              Tel: 520 626 3967
              Fax: 520 621 9288
              Email: rwing@genome.arizona.edu
              PCR Primers
              FORWARD: T7
              BACKWARD: M13r
              Plate: 0108 row: F column: 12
              Seq primer: M13r
              Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..560
              /organism="Zea mays"
              /mol_type="genomic DNA"
              /cultivar="B73"
              /db_xref="taxon:4577"
              /clone="ZMMBBb0108F12"
              /lab_host="DH10B"
              /clone_lib="ZMMBBb"
              /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
              HindIII; Zea mays L. ssp. mays"
ORIGIN

Alignment Scores:
Pred. No.:   151   Length: 560
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x CC740775 (1-560)

Qy   50 SerGlyAlaGlyLeuProSerAlaSerAla 59

ORIGIN

Db   224 TCTGGAGCTGGGCTTCCTAGTGCGTCGGCC 195
             |||||||||||||||||||||||||||||||

RESULT 63
CK691303
LOCUS   CK691303   576 bp   mRNA   linear   EST 30-MAR-2004
DEFINITION   ZF101-P00055-DEPE-F2_O09 GISZF001_ra Danio rerio cDNA clone
              IMAGE:7155107 5', mRNA sequence.
ACCESSION   CK691303
VERSION   CK691303.1 GI:42443639
KEYWORDS   EST.
SOURCE   Danio rerio (zebrafish)
ORGANISM   Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 576)
AUTHORS   Wei,C., Mathavan,S., Thoreau,H., Lim,L., Lee,C. and Ruan,Y.
TITLE   Genome Institute of Singapore, Zebrafish Gene Collection
JOURNAL   Unpublished (2004)
COMMENT   Contact: Ruan Y
              Cloning and Sequencing
              Genome Institute of Singapore
              60 Biopolis Street, #02-01, Genome, Singapore 138672
              Tel: +65 6478 8073
              Fax: +65 6478 9059
              Email: ruanyj@gis.a-star.edu.sg
              GIS Clone ID: ZF101-P00055-BR2_O09
              PCR PRimers
              FORWARD: M13
              BACKWARD: M13
              Plate: ZF101-P00055-BR2 row: O   column: 09
              Seq primer: CCGCATAACTTGTATAGCA
              High quality sequence stop: 576.
FEATURES   Location/Qualifiers
     source   1..576
              /organism="Danio rerio"
              /mol_type="mRNA"
              /strain="Singapore local strain"
              /db_xref="taxon:7955"
              /clone="IMAGE:7155107"
              /tissue_type="Embryo"
              /dev_stage="7 Different embryonic stages(From just
              fertilized Embryos to 72 hours just hatched baby fish)"
              /lab_host="DH10B"
              /clone_lib="GISZF001_ra"
              /note="Vector: pDNR-LIB; Site_1: Sfi A (GGCCATTACGGCC);
              Site_2: Sfi B (GGCCGAGCGGGCC); Priming method: Sfi-(dT)30
              Primed ; Priming sequence:
              5.ATTCTAGAGGCCGAGCGGCGCCGACATG(T)30VN ; Directionally
              cloned, 5' cloning site: Sfi A site GGCCATTACGGCC ; 5'
              linker/adaptor sequence: 5.AAGCAGTGGTATCAACGCAGAGTGGCC ;
              3' cloning site: Sfi B site GGCCGAGCGGGCC ; 3'
              linker/adaptor sequence: same as the priming sequence ;
              Average insert size: 2kb ; For PCR insert analysis: Use
              M13 Forward and reverse primers ; Library Amplified ;
              Recombinants (inserts): 98% ; Library complexity: 5x106 ;
              Full-length construction (method): SMART, a Clontech
              method The pooled tissue RNA was collected and used to
              construct full length enriched cDNA library and also
              served as template to synthesize complex first strand cDNA
              probe. Two high density colony arrays were made from over
              110K cDNA clones and hybridized with the probes. Low
              intensity clones were selected as they represented rare
              expressed clones. The hybridization intensities for all
              clones span from 0 to 1.8 million counts and the low
              abundant class ranged from 0 to 13,000."
ORIGIN

Alignment Scores:
Pred. No.:   155   Length: 576
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0

Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x CK691303 (1-576)

```
Qy    27 SerLeuAlaLeuLeuSerArgProLeuSer 36
         ||||||||||||||||||||||||||||||
Db   351 AGCCTAGCATTGTTATCAAGACCGTTGAGC 380
```

RESULT 64
CR181960/c
LOCUS   CR181960   597 bp   DNA   linear   GSS 06-JUL-2004
DEFINITION   Reverse strand read from insert in 5'HPRT insertion targeting and
               chromosome engineering clone MHPN145i01, genomic survey sequence.
ACCESSION   CR181960
VERSION   CR181960.1 GI:49960809
KEYWORDS   GSS; genome survey sequence; MICER.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 597)
AUTHORS   Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
               Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
               Rogers,J. and Bradley,A.
TITLE   Direct Submission
JOURNAL   Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
               CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES   Location/Qualifiers
     source   1..597
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="MHPN145i01"
                     /clone_lib="MHPN"
ORIGIN

Alignment Scores:
Pred. No.:   Length: 597
Score: 160   Matches: 10
Percent Similarity: 10.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 4.7%   Gaps: 0
11

US-09-989-890-238 (1-212) x CR181960 (1-597)

```
Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   246 GCTGGGCTGCCCTAGTGCCAGCGCTGCTGCG 217
```

RESULT 65
AW300151/c
LOCUS   AW300151   600 bp   mRNA   linear   EST 18-JAN-2000
DEFINITION   xb58b08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2773815 3'
               similar to SW:FKHR_HUMAN Q12778 FORK HEAD DOMAIN PROTEIN FKHR. [1]
               ; mRNA sequence.
ACCESSION   AW300151
VERSION   AW300151.1 GI:6709828
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominidae; Homo.
REFERENCE   1 (bases 1 to 600)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
               Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
               Emmert-Buck, M.D., Ph.D.
               cDNA Library Preparation: M. Bento Soares, Ph.D.
               cDNA Library Arrayed by: Greg Lennon, Ph.D.
               DNA Sequencing by: Washington University Genome Sequencing Center
               Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               www-bio.llnl.gov/bbrp/image/image.html

               Possible reversed clone: polyT not found
               Seq primer: -40UP from Gibco
               High quality sequence stop: 482.
FEATURES   Location/Qualifiers
     source   1..600
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2773815"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Kid11"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.: 161   Length: 600
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 1   Gaps: 0

US-09-989-890-238 (1-212) x AW300151 (1-600)

```
Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   133 GCGGGCCTGCCCTCGGCCTCGGCTGCCGCT 104
```

RESULT 66
CB564205
LOCUS   CB564205   605 bp   DNA   linear   GSS 28-SEP-2003
DEFINITION   tigr-gss-dog-1700032750373B Dog Library Canis familiaris genomic,
               genomic survey sequence.
ACCESSION   CB564205
VERSION   CB564205.1 GI:36880986
KEYWORDS   GSS.
SOURCE   Canis familiaris (dog)
ORGANISM   Canis familiaris
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
               Canis.
REFERENCE   1 (bases 1 to 605)
AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
               Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
               Venter,J.C.
TITLE   The dog genome: survey sequencing and comparative analysis
JOURNAL   Science 301 (5641), 1898-1903 (2003)
PUBMED   14512627
COMMENT   Contact: Kirkness EF
               The Institute for Genomic Research
               Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
               Rockville, MD 20850, USA
               Tel: 301-838-0200
               Fax: 301-838-0208

```
             Email: ekirknes@tigr.org
       Class: shotgun.
FEATURES       Location/Qualifiers
       source  1..605
               /organism="Canis familiaris"
               /mol_type="genomic DNA"
               /strain="Standard Poodle"
               /db_xref="taxon:9615"
               /clone_lib="Dog Library"
               /note="Site 1: BstXI; Libraries were prepared from
               peripheral blood"

ORIGIN

Alignment Scores:
Pred. No.:             162              Length:          605
Score:                 10.00            Matches:         10
Percent Similarity:    100.0%           Conservative:    0
Best Local Similarity: 100.0%           Mismatches:      0
Query Match:           4.7%             Indels:          0
DB:                    10               Gaps:            0

US-09-989-890-238 (1-212) x CE564205 (1-605)

Qy      30 LeuLeuSerArgProLeuSerProProPro 39
           ||||||||||||||||||||||||||||||
Db     276 TTGCTTTCAAGACCCCTCTCCCCCTCCCCCC 305

RESULT 67
BX871243/c
LOCUS       BX871243       618 bp    mRNA      linear    EST 24-MAY-2004
DEFINITION  BX871243 tcbk Oncorhynchus mykiss cDNA clone tcbk0019c.c.19 5prim,
            mRNA sequence.
ACCESSION   BX871243
VERSION     BX871243.2  GI:42785983
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
ORGANISM    Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 618)
AUTHORS     Govoroun,M., Guiguen,Y. and Le Gac,F.
TITLE       Construction and primary characterization of normalized cDNA
            libraries in rainbow trout, Oncorhynchus mykiss
JOURNAL     Unpublished (2003)
COMMENT     On Dec 17, 2003 this sequence version replaced gi:39994830.
            Contact: Guiguen Y
            INRA - SCRIBE
            Campus de beaulieu, RENNES cedex, 35042, France
            Tel: 02.23.48.50.09
            Fax: 02.23.48.50.20
            Email: Yann.Guiguen@beaulieu.rennes.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenaesupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0019 row: c  column: 19
            Seq primer: M13R.
FEATURES       Location/Qualifiers
       source  1..618
               /organism="Oncorhynchus mykiss"
               /mol_type="mRNA"
               /db_xref="taxon:8022"
               /clone="tcbk0019c.c.19"
               /tissue_type="multi-tissues"
               /dev_stage="from embryos to adults"
               /lab_host="DH10B"
               /clone_lib="tcbk"
               /note="Vector: pTT3D-pac; AGENAE Rainbow trout
               distribution - normalized + 2 subtractions; Clone
               distribution : AGENAE Resource centre. Francois PIUMI,
               Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
               Etude du genome (LREG), Domaine de Vilvert, 78352,
               Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
```

```
                                          (0) 1.34.65.22.73"

ORIGIN

Alignment Scores:
Pred. No.:             166              Length:          618
Score:                 10.00            Matches:         10
Percent Similarity:    100.0%           Conservative:    0
Best Local Similarity: 100.0%           Mismatches:      0
Query Match:           4.7%             Indels:          0
DB:                    5                Gaps:            0

US-09-989-890-238 (1-212) x BX871243 (1-618)

Qy      30 LeuLeuSerArgProLeuSerProProPro 39
           ||||||||||||||||||||||||||||||
Db     122 TTACTCTCCCGGCCCCTCTCTCCCCCTCCA 93

RESULT 68
BB659723/c
LOCUS       BB659723       620 bp    mRNA      linear    EST 26-OCT-2001
DEFINITION  BB659723 RIKEN full-length enriched, 13 days embryo lung Mus
            musculus cDNA clone D430004P20 5', mRNA sequence.
ACCESSION   BB659723
VERSION     BB659723.1  GI:16493544
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 620)
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,N., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Saito,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
            RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
            Unpublished (2001)
TITLE       Contact: Yoshihide Hayashizaki
JOURNAL     Laboratory for Genome Exploration Research Group, RIKEN Genomic
COMMENT     Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu.M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res.   10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res.  11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
```

```
FEATURES     Location/Qualifiers
source       1..620
             /organism="Mus musculus"
             /mol_type="mRNA"
             /db_xref="taxon:10090"
             /clone="D430004P20"
             /tissue_type="lung"
             /dev_stage="13 days embryo"
             /lab_host="DH10B"
             /clone_lib="RIKEN full-length enriched, 13 days embryo
             lung"
             /note="Site_1: SalI; Site_2: BamHI; cDNA library was
             prepared and sequenced in Mouse Genome Encyclopedia
             Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in
             RIKEN. Division of Experimental Animal Research in Riken
             contributed to prepare mouse tissues. 1st strand cDNA was
             primed with a primer [5'
             GAGAGAGAGAGCGGCCGCCAACTCGAGTTTTTTTTTTTTTVN 3'], cDNA was
             prepared by using trehalose thermo-activated reverse
             transcriptase and subsequently enriched for full-length by
             cap-trapper. Second strand cDNA was prepared with the
             primer adapter of sequence [5'
             GAGAGAGAGGATTCGGAGTTAATTAATTAATCCCCCCCCCCCCCC 3']. cDNA
             was cleaved with BamHI and XhoI. Vector: a modified
             pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Alignment Scores:
Pred. No.:               166          Length:        620
Score:                   10.00        Matches:       10
Percent Similarity:      100.0%       Conservative:  0
Best Local Similarity:   100.0%       Mismatches:    0
Query Match:             4.7%         Indels:        0
DB:                                   Gaps:          0

US-09-989-890-238 (1-212) x BB659723 (1-620)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAlaAla 61
        |||||||||||||||||||||||||||||||||
Db  160 GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 131

RESULT 69
BI824376
LOCUS       BI824376        647 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  603038602F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179344 5',
            mRNA sequence.
ACCESSION   BI824376
VERSION     BI824376.1  GI:15935926
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 647)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            cDNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11447 row: d column: 01
            High quality sequence stop: 644.

FEATURES     Location/Qualifiers
source       1..647
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:5179344"
             /lab_host="DH10B"
             /note="Organ: pooled brain, lung, testis; Vector:
             pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
             source anonymous pool of 6 male brains, age range 23-27; 1
             male lung, age 27; and 1 male testis, age 69. Library is
             oligo-dT primed and directionally cloned (EcoRV site is
             destroyed upon cloning). Average insert size is 1.8 kb,
             insert size range 1-3 kb. Library is normalized and
             enriched for full-length clones and was constructed by C.
             Gruber (Invitrogen). Research Genetics tracking code
             021. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:               173          Length:        647
Score:                   10.00        Matches:       10
Percent Similarity:      100.0%       Conservative:  0
Best Local Similarity:   100.0%       Mismatches:    0
Query Match:             4.7%         Indels:        0
DB:                                   Gaps:          0

US-09-989-890-238 (1-212) x BI824376 (1-647)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
        ||||||||||||||||||||||||||||||
Db  524 GCGGGCCTGCCTCCGCCCTCGGCTGCCGCT 553

RESULT 70
BB393441
LOCUS       BB393441        648 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  BB393441 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C230083P08 3', mRNA sequence.
ACCESSION   BB393441
VERSION     BB393441.2  GI:16409146
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 648)
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL     Unpublished (2001)
COMMENT     On Jul 13, 2000 this sequence version replaced gi:9116416.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,Y., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
```

sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res.. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., and
Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

FEATURES Location/Qualifiers
 source 1..648
    /organism="Mus musculus"
    /mol_type="mRNA"
    /db_xref="taxon:10090"
    /clone="C230083P08"
    /tissue_type="cerebellum"
    /dev_stage="0 day neonate"
    /clone_lib="RIKEN full-length enriched, 0 day neonate
    cerebellum"
    /lab_host="DH10B"
    /note="Site 1: SalI; Site 2: BamHI; cDNA library was
    prepared by using trehalose thermo-activated reverse
    transcriptase and subsequently enriched for full-length by
    cap-trapper. cDNA went through one round of normalization
    to Rot = 20.0 and subtraction to Rot = 479.0. Second
    strand cDNA was prepared with the primer adapter of
    sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC
    3']. cDNA was cleaved with XhoI and BamHI. Vector: a
    modified pBluescript KS(+) after bulk excision from Lambda
    FLC I."

ORIGIN

Alignment Scores:
Pred. No.:   173   Length: 648
Score: 10.00   Matches: 10
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x BB393441 (1-648)

Qy 46 ProGlyCysGlySerGlyAlaGlyLeuPro 55
   |||||||||||||||||||||||||||||
Db 209 CCTGGTGTGGCAGTGGTCCAGGGCTTCCT 238

RESULT 71
CD598528
LOCUS   CD598528   650 bp mRNA linear EST 16-JUN-2003
DEFINITION RK114A2E10.T3 Zebrafish Kidney Marrow cDNA library Danio rerio cDNA
    clone RK114A2E10 5', mRNA sequence.
ACCESSION CD598528
VERSION   CD598528.1 GI:31779880
KEYWORDS EST.
SOURCE   Danio rerio (zebrafish)
 ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 650)
AUTHORS   Song,H.D., Sun,X.J., Deng,M., Zhang,G.W., Zhou,Y., Wu,X.Y.,
    Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y., Zon,L.I.,
    Kanki,J.P., Liu,T.X., Look,A.T. and Chen,Z.
TITLE   Hematopoietic gene expression profile in zebrafish kidney marrow
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (46), 16240-16245 (2004)
PUBMED 15520368
COMMENT Contact: Chen Z.
    State Key Lab for Medical Genomics
    Shanghai Institute of Hematology, Ruijin Hospital Affiliated to
    Shanghai Second Medical University
    197 Rui Jin Road II, Shanghai 200025, P. R. China
    Tel: 86-21-64740490
    Fax: 86-21-64743206
    Email: zchen@stn.sh.cn
    Seq primer: T3.

FEATURES Location/Qualifiers
 source 1..650
    /organism="Danio rerio"
    /mol_type="mRNA"
    /db_xref="taxon:7955"
    /clone="RK114A2E10"
    /dev_stage="mature"
    /clone_lib="Zebrafish Kidney Marrow cDNA library"
    /note="Organ: kidney; Vector: pBS-CMV; Site 1: XhoI;
    Site_2: EcoRI; Total RNA was extracted from the kidney
    tissues of mature zebrafish. The poly (A)+ RNA fraction
    was separated from total RNA by oligo (dT) cellulose
    chromatography. Library was initially constructed in the
    lambdaZAP Express vector (Stratagene) and in vivo excised
    into pBS-CMV vector."

ORIGIN

Alignment Scores:
Pred. No.:   174   Length: 650
Score: 10.00   Matches: 10
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 6

US-09-989-890-238 (1-212) x CD598528 (1-650)

Qy 27 SerLeuAlaLeuLeuSerArgProLeuSer 36
   ||||||||||||||||||||||||||||||
Db 311 AGCCTAGCATTGTTATCAAGACCGGTGAGC 340

RESULT 72
BM265585
LOCUS   BM265585   652 bp mRNA linear EST 18-DEC-2001
DEFINITION fw57d06.y1 Sugano SJD adult male Danio rerio cDNA clone
    IMAGE:5611811 5' similar to TR:Q9Y4E0 Q9Y4E0 KIAA0595 PROTEIN ;,
    mRNA sequence.
ACCESSION BM265585
VERSION   BM265585.1 GI:17928625
KEYWORDS EST.
SOURCE   Danio rerio (zebrafish)
 ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 652)
AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,F., Marra,M.,
    Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
    Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
    Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
    Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
    Waterston,R. and Wilson,R.
TITLE   WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)

COMMENT   Other ESTs: fw5d06.x1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution information can be found through the I.M.A.G.E.
            Consortium/LLNL, send email to: info@image.llnl.gov
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 522.
            Heubnerweg 6, D-14059 Berlin, Germany
            Email: www.rzpd.de
            RZPD; RZPDp9016D1639.
            RZPDLIB; (Human T-Lymphocytes) RZPD LIB No.9016
            http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=9016 Contact:
            Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 100
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available from RZPD;
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPDp9016D1639
            contact RZPD (product-support@rzpd.de) for further information.
            Primer name: qe3_4 , Primer sequence: CGGATAACAATTCACACAG.

FEATURES   Location/Qualifiers
     source   1..652
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:5611811"
                     /sex="male"
                     /tissue_type="whole body"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /clone_lib="Sugano SJD adult male"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACCATGTG);
                     Site_2: DraIII (CACTGTGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [ATGTGGCCCTTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed and donated by Dr. Sumio
                     Sugano (University of Tokyo Institute of Medical Science).
                     Custom primers for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA. "

FEATURES   Location/Qualifiers
     source   1..659
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="RZPDp9016D1639"
                     /tissue_type="T-Lymphocytes"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RZPD no.9016"
                     /note="Vector: pQE80LSN_cloned; Site_1: SalI; Site_2:
                     NotI; vector:
                     http://www.rzpd.de/info/vectors/pQE80LSN_cloned_pic.shtml
                     ; 1st strand cDNA was prepared from mRNA obtained from
                     human T-Lymphocytes with a NotI - oligo(dT) primer (5'
                     GACTAGTTCTAGATCGGCAGGCGGCCGCCGCCTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to SalI adaptors,
                     digested with NotI and cloned into the NotI and SalI sites
                     of the pQE80LSN_cloned vector"

ORIGIN

Alignment Scores:
Pred. No.:   174   Length: 652
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x BM265585 (1-652)

Qy   27 SerLeuAlaLeuLeuSerArgProLeuSer 36
        |||||||||||||||||||||||||||||||
Db 441 AGCCTAGCATTGTTATCAAGACCGTTGAGC 470

ORIGIN

Alignment Scores:
Pred. No.:   176   Length: 659
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 7   Gaps: 0

US-09-989-890-238 (1-212) x CR976136 (1-659)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
        ||||||||||||||||||||||||||||||
Db 552 GCGGGCCTGCCCTCGGCCTCGGCTGCCGCT 581

RESULT 73
CR976136
LOCUS   CR976136 RZPD no.9016 Homo sapiens cDNA clone RZPDp9016D1639 5',   659 bp   mRNA   linear   EST 22-JUN-2005
DEFINITION CR976136 RZPD no.9016 Homo sapiens cDNA clone RZPDp9016D1639 5',
            mRNA sequence.
ACCESSION   CR976136
VERSION   CR976136.1 GI:68146090
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 659)
  AUTHORS   Heil,O., Ebert,L., Hennig,S., Henze,S., Radelof,U., Schneider,D.
            and Korn,B.
  TITLE   Human T-Lymphocytes library
  JOURNAL   Unpublished (2005)
COMMENT   Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

RESULT 74
BM051939/c
LOCUS   BM051939   668 bp   mRNA   linear   EST 07-NOV-2001
DEFINITION 603638841F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5419690 5',
            mRNA sequence.
ACCESSION   BM051939
VERSION   BM051939.1 GI:16781206
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1 (bases 1 to 668)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.

```
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1873 row: b column: 11
High quality sequence stop: 664.
FEATURES
  source          Location/Qualifiers
                  1..668
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:5419690"
                  /tissue_type="Burkitt lymphoma"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NIH_MGC_8"
                  /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                  EcoRI; cDNA made by oligo-dT priming. Directionally
                  cloned into EcoRI/XhoI sites using the following 5'
                  adaptor: GGCACGAG(G). Size-selected >500bp for average
                  insert size 1.8kb. Library constructed by Ling Hong in
                  the laboratory of Gerald M. Rubin (University of
                  California, Berkeley) using ZAP-cDNA synthesis kit
                  (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:                    178          Length:          668
Score:                        10.00        Matches:         10
Percent Similarity:           100.0%       Conservative:    0
Best Local Similarity:        100.0%       Mismatches:      0
Query Match:                  4.7%         Indels:          0
DB:                           3            Gaps:            0

US-09-989-890-238 (1-212) x BM051939 (1-668)

Qy     26 SerSerLeuAlaLeuSerArgProLeu 35
          ||||||||||||||||||||||||
Db    410 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 381

RESULT 75
BM051456/c
LOCUS      BM051456     686 bp    mRNA    linear    EST 07-NOV-2001
DEFINITION 603638116F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5419610 5',
           mRNA sequence.
ACCESSION  BM051456
VERSION    BM051456.1  GI:16780723
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Hominidae; Homo.
REFERENCE  1  (bases 1 to 686)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1873 row: o column: 03
           High quality sequence stop: 685.
FEATURES
  source          Location/Qualifiers
                  1..686
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:5419610"
                  /tissue_type="Burkitt lymphoma"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NIH_MGC_8"
                  /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                  EcoRI; cDNA made by oligo-dT priming. Directionally
                  cloned into EcoRI/XhoI sites using the following 5'
                  adaptor: GGCACGAG(G). Size-selected >500bp for average
                  insert size 1.8kb. Library constructed by Ling Hong in
                  the laboratory of Gerald M. Rubin (University of
                  California, Berkeley) using ZAP-cDNA synthesis kit
                  (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:                    183          Length:          686
Score:                        10.00        Matches:         10
Percent Similarity:           100.0%       Conservative:    0
Best Local Similarity:        100.0%       Mismatches:      0
Query Match:                  4.7%         Indels:          0
DB:                           3            Gaps:            0

US-09-989-890-238 (1-212) x BM051456 (1-686)

Qy     26 SerSerLeuAlaLeuSerArgProLeu 35
          ||||||||||||||||||||||||
Db    411 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 382

RESULT 76
DR001418
LOCUS      DR001418     695 bp    mRNA    linear    EST 17-MAY-2005
DEFINITION TC118890 Human fetal brain, large insert, pCMV expression library
           Homo sapiens cDNA clone TC118890 5' similar to Homo sapiens
           forkhead box O1A (rhabdomyosarcoma) (FOXO1A), mRNA sequence.
ACCESSION  DR001418
VERSION    DR001418.1  GI:66261291
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Hominidae; Homo.
REFERENCE  1  (bases 1 to 695)
AUTHORS    Birkett,C., Cho,J., Gau,Y., Hamer,R., Kelly,S., Kovacs,K., Liu,L.,
           Liu,X., Porter,J., Sachs,A., Shu,Y., Sun,Z., Wong,J., Wu,M.,
           Zhang,X., Jay,G. and He,W.
TITLE      High-throughput cloning of full-length human cDNAs directly from
           cDNA libraries optimized for large and rare transcripts
JOURNAL    Unpublished (2005)
COMMENT    Contact: Kovacs, KF
           High Throughput cDNA Cloning
           OriGene Technologies, Inc. ( www.origene.com )
           6 Taft Court, Suite 100, Rockville, MD 20850, USA
           Tel: 301 340 3188
           Fax: 301 340 8606
           Email: cDNA@origene.com
           This EST submission is part of an on-going human full-length
           cloning project at OriGene Technologies, Inc.
           Please contact OriGene for access.
           OriGene Technologies, Inc.
           6 Taft Ct. Suite 100
           Rockville, MD 20850
           Tel: (301) 340-3188
           http://www.origene.com
           Seq primer: pCMV6 5prime forward vector primer, OriGene
           Technologies Inc.
FEATURES
  source          Location/Qualifiers
                  1..695
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="TC118890"
                  /tissue_type="Fetal Brain"
                  /clone_lib="Human fetal brain, large insert, pCMV
                  expression library"
```

```
                     /note="Organ: Fetal Brain; Vector: pCMV6-XL4; Site 1:
                     EcoR1; Site_2: XhoI/SalI compatible end ligatio; Oligo-dT
                     primed reverse transcription optimized for large and GC
                     rich mRNA transcripts, cDNA size selection, optimized
                     ligation for large inserts into mammalian expression
                     vector, random clones selected for end sequence
                     verification of full-length genes"

ORIGIN

Alignment Scores:
Pred. No.:                185         Length:        695
Score:                    10.00       Matches:       10
Percent Similarity:       100.0%      Conservative:  0
Best Local Similarity:    100.0%      Mismatches:    0
Query Match:              4.7%        Indels:        0
DB:                                   Gaps:          0

US-09-989-890-238 (1-212) x DR001418 (1-695)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   578 GCGGGCCTGCCCTCGGCCTCGGCTGCCGCT 607

RESULT 77
CF727607/c
LOCUS       CF727607                702 bp    mRNA     linear   EST 09-OCT-2003
DEFINITION  UI-M-HB0-cki-h-12-0-UI.r1 NIH_BMAP_HB0 Mus musculus cDNA clone
            IMAGE:3054835 5', mRNA sequence.
ACCESSION   CF727607
VERSION     CF727607.1  GI:37601775
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 702)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..702
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3054835"
                     /tissue_type="whole eye"
                     /dev_stage="embryo 12.5,13.5,14.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_HB0"
                     /note="Organ: Eye; Vector: pYX- Asc; Site 1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     to mRNA size fraction. Double strand cDNA was
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site .Double strand cDNA was
                     size selected according to mRNA size fraction ,ligated
                     with EcoR I adaptor ,digested with NotI and then cloned
                     directionally into pYX-Asc vector . The library tag
                     sequence located between the Not I site and the polyA tail
                     is TTATTGAAGT. This library was created for the University
```

```
                     Iowa Brain Anatomy Project (BMAP) ; 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:
Pred. No.:                187         Length:        702
Score:                    10.00       Matches:       10
Percent Similarity:       100.0%      Conservative:  0
Best Local Similarity:    100.0%      Mismatches:    0
Query Match:              4.7%        Indels:        0
DB:                                   Gaps:          0

US-09-989-890-238 (1-212) x CF727607 (1-702)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   362 GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 333

RESULT 78
BX890304/c
LOCUS       BX890304                705 bp    mRNA     linear   EST 27-JUL-2004
DEFINITION  BX890304 tcbk Oncorhynchus mykiss cDNA clone tcbk0033c.k.12 5prim,
            mRNA sequence.
ACCESSION   BX890304
VERSION     BX890304.2  GI:43411889
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
ORGANISM    Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 705)
AUTHORS     Govoroun,M., Guiguen,Y. and Le Gac,F.
TITLE       Construction and primary characterization of normalized cDNA
            libraries in rainbow trout, Oncorhynchus mykiss
JOURNAL     Unpublished (2003)
COMMENT     On Dec 18, 2003 this sequence version replaced gi:40140704.
            Contact: Guiguen Y
            INRA - SCRIBE
            Campus de beaulieu, RENNES cedex, 35042, France
            Tel: 02.23.48.50.09
            Fax: 02.23.48.50.20
            Email: Yann.Guiguen@beaulieu.rennes.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0033   row: k  column: 12
            Seq primer: M13R.
FEATURES             Location/Qualifiers
     source          1..705
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="tcbk0033c.k.12"
                     /tissue_type="multi-tissues"
                     /dev_stage="from embryos to adults"
                     /lab_host="DH10B"
                     /clone_lib="tcbk"
                     /note="Vector: pT7T3D-pac; AGENAE Rainbow trout
                     multi-tissues - normalized + 2 subtractions; Clone
                     distribution : AGENAE Resource centre. Francois PIUMI,
                     Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
                     Etude du genome (LREG), Domaine de Vilvert, 78352,
                     Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
                     (0) 1.34.65.22.73"

ORIGIN

Alignment Scores:
Pred. No.:                188         Length:        705
Score:                    10.00       Matches:       10
Percent Similarity:       100.0%      Conservative:  0
Best Local Similarity:    100.0%      Mismatches:    0
```

Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x BX890304 (1-705)

Qy   30 LeuLeuSerArgProLeuSerProProPro 39
           ||||||||||||||||||||||||||||||
Db   122 TTACTCTCCGGCCCTGTCTCTCCCCCTCCA 93

RESULT 79
CB246007/c
LOCUS   CB246007   710 bp   mRNA   linear   EST 09-JUL-2003
DEFINITION   UI-M-FO0-cdu-n-21-0-UI.r1 NIH_BMAP_FO0 Mus musculus cDNA clone
             IMAGE:6834742 5', mRNA sequence.
ACCESSION   CB246007
VERSION   CB246007.1 GI:28367651
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 710)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
             (BMAP)
             Seq primer: pYX-5.
FEATURES   Location/Qualifiers
     source   1..710
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6834742"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FO0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     BonaIdo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is TGAGAGAGCC. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."

ORIGIN

Alignment Scores:
Pred. No.:   189   Length: 710
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x CB246007 (1-710)

Qy   52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||||||||
Db   301 GCTGGGCTGCCTAGTGCCAGCGCTGCTGCG 272

RESULT 80
DR608449
LOCUS   DR608449   711 bp   mRNA   linear   EST 11-JUL-2005
DEFINITION   EST998577 FvG Gibberella moniliformis cDNA clone FVGCK02, mRNA
             sequence.
ACCESSION   DR608449
VERSION   DR608449.1 GI:70683097
KEYWORDS   EST.
SOURCE   Gibberella moniliformis
ORGANISM   Gibberella moniliformis
             Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
             Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE   1 (bases 1 to 711)
AUTHORS   Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L, Lee,Y.,
             Utterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D.,
             Kendra,D.F., Town,C.D. and Whitelaw,C.A.
TITLE   Analysis of 87,000 expressed sequence tags reveals alternatively
             spliced introns in multiple genes of the fumonisin gene cluster
JOURNAL   Unpublished (2005)
COMMENT   Contact: Brown, D.W.
             USDA/ARS/NCAUR
             USDA
             1815 N. University St, Peoria, IL 61604, USA
             Tel: 309 681 6230
             Fax: 309 681 6689
             Email: browndw@ncaur.usda.gov
             TIGR sequence name: FVGCK02TH
             Seq primer: AAT TAA CCC TCA CTA AAG GG.
FEATURES   Location/Qualifiers
     source   1..711
                     /organism="Gibberella moniliformis"
                     /mol_type="mRNA"
                     /strain="m3125"
                     /db_xref="taxon:117187"
                     /clone="FVGCK02"
                     /tissue_type="mycelia"
                     /clone_lib="FvG"
                     /note=vector: pBlueScript II SK(+) XR; Site_1: EcoRI;
                     Site_2: XhoI; anamorph: Fusarium verticillioides. Mycelia
                     was collected after growth in liquid GYAM medium for 96
                     hours. Cultures were vacuum filtered and the mycelial mats
                     were frozen in liquid nitrogen, ground to a powder, and
                     then added to TRIzol Reagent (Invitrogen, Carlsbad CA) at
                     approximately 1 g mycelia per 10 ml TRIzol. The cDNA was
                     directionally ligated into the pBlueScript II SK(+) XR
                     vector (cDNA Synthesis Kit; Stratagene)."

ORIGIN

Alignment Scores:
Pred. No.:   189   Length: 711
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x DR608449 (1-711)

Qy   30 LeuLeuSerArgProLeuSerProProPro 39
           ||||||||||||||||||||||||||||||
Db   603 CTCCTTTCAAGACCACTTTCGCCGCCACCT 632

RESULT 81
DR614353
LOCUS   DR614353   714 bp   mRNA   linear   EST 11-JUL-2005
DEFINITION   EST1004481 FvH Gibberella moniliformis cDNA clone FVHBF85, mRNA

ACCESSION   DR614353
VERSION   DR614353.1   GI:70689001
KEYWORDS   EST.
SOURCE   Gibberella moniliformis
ORGANISM   Gibberella moniliformis
             Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
             Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE   1 (bases 1 to 714)
AUTHORS   Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L, Lee,Y.,
             Utterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D.,
             Kendra,D.F., Town,C.D. and Whitelaw,C.A.
TITLE   Analysis of 87,000 expressed sequence tags reveals alternatively
             spliced introns in multiple genes of the fumonisin gene cluster
JOURNAL   Unpublished (2005)
COMMENT   Contact: Brown, D.W.
             USDA/ARS/NCAUR
             USDA
             1815 N. University St, Peoria, IL 61604, USA
             Tel: 309 681 6230
             Fax: 309 681 6689
             Email: browndw@ncaur.usda.gov
             TIGR sequence name: FVHBF85TH
             Seq primer: AAT TAA CCC TCA CTA AAG GG.
FEATURES   Location/Qualifiers
     source   1..714
                     /organism="Gibberella moniliformis"
                     /mol_type="mRNA"
                     /strain="m3125"
                     /db_xref="taxon:117187"
                     /clone="FVHBF85"
                     /tissue_type="mycelia"
                     /clone_lib="FvH"
                     /note="Vector: pBlueScript II SK(+) XR; Site_1: EcoRI;
                     Site_2: XhoI; anamorph: Fusarium verticillioides. Library
                     FvH was prepared from germinating conidia in an aqueous
                     extract of maize seedlings. The aqueous extract was
                     obtained by germinating surface-sterilized maize seeds for
                     four days on 1% water agar supplemented with 100 ug/ml
                     streptomycin sulfate. Under sterile conditions, 70 g of
                     excised roots and shoots from the resulting seedlings was
                     blended with 250 ml water, filtered once through Miracloth
                     (Calbiochem) and then again through #1 Whatman filter
                     paper. The extracts were inoculated to a spore
                     concentration of 5 x 106 conidia per ml and then incubated
                     for 10 hours at 30C with shaking at 200 RPM. Cultures were
                     vacuum filtered and the mycelial mats were frozen in
                     liquid nitrogen, ground to a powder, and then added to
                     TRIzol Reagent (Invitrogen, Carlsbad CA) at approximately
                     1 g mycelia per 10 ml TRIzol. The cDNA was directionally
                     ligated into the pBlueScript II SK(+) XR vector (cDNA
                     Synthesis Kit; Stratagene)."
ORIGIN

Alignment Scores:
Pred. No.:   190   Length: 714
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x DR614353 (1-714)

Qy   30 LeuLeuSerArgProLeuSerProPro 39
           |||||||||||||||||||||||||||
Db   374 CTCCTTCAAGACCACTTTCGCCGCCACCT 403

RESULT 82
CC437175/c
LOCUS   CC437175   714 bp   DNA   linear   GSS 20-MAY-2003
DEFINITION   PUHBS30TB ZM_0.6_1.0.KB Zea mays genomic clone ZMMBTa415E11,
             genomic survey sequence.

ACCESSION   CC437175
VERSION   CC437175.1   GI:30935023
KEYWORDS   GSS.
SOURCE   Zea mays
ORGANISM   Zea mays
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
             clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 714)
AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
             Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
             Bennetzen,J.
TITLE   Maize Genomics Consortium
JOURNAL   Unpublished (2003)
COMMENT   Other_GSSs: PUHBS30TD
             Contact: Cathy Whitelaw
             TIGR
             9712 Medical Center Drive, Rockville, MD 20850, USA
             Tel: 301-838-5843
             Fax: 301-838-0208
             Email: whitelaw@tigr.org
             Seq primer: TR
             Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..714
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa415E11"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:   190   Length: 714
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 9   Gaps: 0

US-09-989-890-238 (1-212) x CC437175 (1-714)

Qy   50 SerGlyAlaGlyLeuProSerAlaSerAla 59
           ||||||||||||||||||||||||||||||
Db   295 TCTGGAGCTGGGCTTCCTAGTGCGTCGGCC 266

RESULT 83
BH353352
LOCUS   CH230-68M10.TV CHORI-230   721 bp   DNA   linear   GSS 03-DEC-2001
DEFINITION   CH230-68M10, genomic survey sequence.

ACCESSION   BH353352
VERSION   BH353352.1   GI:17284086
KEYWORDS   GSS.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 721)
AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
             Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
             Riggs,F., de Jong,P. and Fraser,C.M.
TITLE   Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL   Unpublished (1999)
COMMENT   Other_GSSs: CH230-68M10.TJ
             Contact: Shaying Zhao
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA

```
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 68  row: M  column: 10
Seq primer: T7
Class: BAC ends.
FEATURES    Location/Qualifiers
source      1..721
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /strain="BN/SsNHsd/MCW"
            /db_xref="taxon:10116"
            /clone="CH230-68M10"
            /sex="Female"
            /cell_type="Brain"
            /clone_lib="CHORI-230 Segment 1"
            /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
            CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
            Pieter de Jong"

ORIGIN

Alignment Scores:
Pred. No.:              192           Length:            721
Score:                  10.00         Matches:           10
Percent Similarity:     100.0%        Conservative:      0
Best Local Similarity:  100.0%        Mismatches:        0
Query Match:            4.7%          Indels:            0
DB:                     9             Gaps:              0

US-09-989-890-238 (1-212) x BH353352 (1-721)

Qy      25 ProSerSerLeuAlaLeuLeuSerArgPro 34
           ||||||||||||||||||||||||||||||
Db     235 CCATCCAGCTTAGCTCTTCTCTCCAGACCC 264

RESULT 84
BX913421/c
LOCUS       BX913421              724 bp    mRNA    linear   EST 27-JUL-2004
DEFINITION  BX913421 tcbk Oncorhynchus mykiss cDNA clone tcbk0074c.i.19 5prim,
            mRNA sequence.
ACCESSION   BX913421
VERSION     BX913421.2  GI:43422650
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   Govoroun,M., Guiguen,Y. and Le Gac,F.
  TITLE     Construction and primary characterization of normalized cDNA
            libraries in rainbow trout, Oncorhynchus mykiss
  JOURNAL   Unpublished (2003)
COMMENT     On Jan 22, 2004 this sequence version replaced gi:41130229.
            Contact: Guiguen Y
            INRA - SCRIBE
            Campus de beaulieu, RENNES cedex, 35042, France
            Tel: 02.23.48.50.09
            Fax: 02.23.48.50.20
            Email: Yann.Guiguen@beaulieu.rennes.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at signasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0074  row: i  column: 19
            Seq primer: M13R.
FEATURES    Location/Qualifiers
source      1 ..724
```

```
            /organism="Oncorhynchus mykiss"
            /mol_type="mRNA"
            /db_xref="taxon:8022"
            /clone="tcbk0074c.i.19"
            /tissue_type="multi-tissues"
            /dev_stage="from embryos to adults"
            /lab_host="DH10B"
            /clone_lib="tcbk"
            /note="Vector: pT7T3D-pac; AGENAE Rainbow trout
            multi-tissues - normalized + 2 subtractions; Clone
            distribution: AGENAE Resource centre. Francois PIUMI,
            Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
            Etude du genome (LREG), Domaine de Vilvert, 78352,
            Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
            (0) 1.34.65.22.73"

ORIGIN

Alignment Scores:
Pred. No.:              192           Length:            724
Score:                  10.00         Matches:           10
Percent Similarity:     100.0%        Conservative:      0
Best Local Similarity:  100.0%        Mismatches:        0
Query Match:            4.7%          Indels:            0
DB:                     5             Gaps:              0

US-09-989-890-238 (1-212) x BX913421 (1-724)

Qy      30 LeuLeuSerArgProLeuSerProProPro 39
           ||||||||||||||||||||||||||||||
Db     122 TTACTCTCCGCCCTTCTCTTCTCCCCCTCCA 93

RESULT 85
DR605627
LOCUS       DR605627              724 bp    mRNA    linear   EST 11-JUL-2005
DEFINITION  EST995755 FvG Gibberella moniliformis cDNA clone FVGBC18, mRNA
            sequence.
ACCESSION   DR605627
VERSION     DR605627.1  GI:70680275
KEYWORDS    EST.
SOURCE      Gibberella moniliformis
  ORGANISM  Gibberella moniliformis
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L, Lee,Y.,
            Utterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D.,
            Kendra,D.F., Town,C.D. and Whitelaw,C.A.
  TITLE     Analysis of 87,000 expressed sequence tags reveals alternatively
            spliced introns in multiple genes of the fumonisin gene cluster
  JOURNAL   Unpublished (2005)
            Contact: Brown, D.W.
            USDA/ARS/NCAUR
            USDA
            1815 N. University St, Peoria, IL 61604, USA
            Tel: 309 681 6230
            Fax: 309 681 6689
            Email: browndw@ncaur.usda.gov
            TIGR sequence name: FVGBC18TH
            Seq primer: AAT TAA CCC TCA CTA AAG GG.
FEATURES    Location/Qualifiers
source      1 ..724
            /organism="Gibberella moniliformis"
            /mol_type="mRNA"
            /strain="m3125"
            /db_xref="taxon:117187"
            /clone="FVGBC18"
            /tissue_type="mycelia"
            /clone_lib="FvG"
            /note="Vector: pBlueScript II SK(+) XR, Site1: EcoRI;
            Site_2: XhoI; anamorph: Fusarium verticillioides. Mycelia
            was collected after growth in liquid GYAM medium for 96
            hours. Cultures were vacuum filtered and the mycelial mats
            were frozen in liquid nitrogen, ground to a powder, and
```

```
            then added to TRIzol Reagent (Invitrogen, Carlsbad CA) at
            approximately 1 g mycelia per 10 ml TRIzol. The cDNA was
            directionally ligated into the pBluescript II SK(+) XR
            vector (cDNA Synthesis Kit, Stratagene)."
ORIGIN

Alignment Scores:
Pred. No.:              192            Length:          724
Score:                  10.00          Matches:         10
Percent Similarity:     100.0%         Conservative:    0
Best Local Similarity:  100.0%         Mismatches:      0
Query Match:            4.7%           Indels:          0
DB:                     8              Gaps:            0

US-09-989-890-238 (1-212) x DR605627 (1-724)

Qy   30  LeuLeuSerArgProLeuSerProProPro 39
         ||||||||||||||||||||||||||
Db  671  CTCCTTTCAAGACCACTTTCGCCGCCACCT 700


RESULT 86
CC437178
LOCUS       CC437178    724 bp    DNA    linear   GSS 20-MAY-2003
DEFINITION  PUHBS30TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa415B11,
            genomic survey sequence.
ACCESSION   CC437178
VERSION     CC437178.1  GI:30935027
KEYWORDS    GSS.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 724)
AUTHORS     Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
TITLE       Maize Genomics Consortium
JOURNAL     Unpublished (2003)
COMMENT     Other GSSs: PUHBS30TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES    Location/Qualifiers
source      1..724
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBTa415B11"
            /clone_lib="ZM_0.6_1.0_KB"
            /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
            CoT selected genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:              192            Length:          724
Score:                  10.00          Matches:         10
Percent Similarity:     100.0%         Conservative:    0
Best Local Similarity:  100.0%         Mismatches:      0
Query Match:            4.7%           Indels:          0
DB:                     9              Gaps:            0

US-09-989-890-238 (1-212) x CC437178 (1-724)

Qy   50  SerGlyAlaGlyLeuProSerAlaSerAla 59
         ||||||||||||||||||||||||||
Db  428  TCTCGAGCTGGGCTTCCTAGTGCGTCGGCC 457
```

```
RESULT 87
BX912624/c
LOCUS       BX912624    tcbk Oncorhynchus mykiss cDNA clone tcbk0072c.f.19 5prim,
                        726 bp    mRNA    linear   EST 27-JUL-2004
DEFINITION  BX912624 tcbk Oncorhynchus mykiss cDNA clone tcbk0072c.f.19 5prim,
            mRNA sequence.
ACCESSION   BX912624
VERSION     BX912624.2  GI:4342058S
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
ORGANISM    Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1 (bases 1 to 726)
AUTHORS     Govoroun,M., Guiguen,Y. and Le Gac,F.
TITLE       Construction and primary characterization of normalized cDNA
            libraries in rainbow trout, Oncorhynchus mykiss
JOURNAL     Unpublished (2003)
COMMENT     On Jan 22, 2004 this sequence version replaced gi:41129432.
            Contact: Guiguen Y
            INRA - SCRIBE
            Campus de beaulieu, RENNES cedex, 35042, France
            Tel: 02.23.48.50.09
            Fax: 02.23.48.50.20
            Email: Yann.Guiguen@beaulieu.rennes.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0072  row: f  column: 19
            Seq primer: M13R.
FEATURES    Location/Qualifiers
source      1..726
            /organism="Oncorhynchus mykiss"
            /mol_type="mRNA"
            /db_xref="taxon:8022"
            /clone="tcbk0072c.f.19"
            /tissue_type="multi-tissues"
            /dev_stage="from embryos to adults"
            /lab_host="DH10B"
            /clone_lib="tcbk"
            /note="Vector: pT7T3D-pac; AGENAE Rainbow trout
            multi-tissues - normalized + 2 subtractions; Clone
            distribution - AGENAE Resource centre. Francois PIUMI,
            Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
            Etude du genome (LREG), Domaine de Vilvert, 78352,
            Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
            (0) 1.34.65.22.73"
ORIGIN

Alignment Scores:
Pred. No.:              193            Length:          726
Score:                  10.00          Matches:         10
Percent Similarity:     100.0%         Conservative:    0
Best Local Similarity:  100.0%         Mismatches:      0
Query Match:            4.7%           Indels:          0
DB:                     5              Gaps:            0

US-09-989-890-238 (1-212) x BX912624 (1-726)

Qy   30  LeuLeuSerArgProLeuSerProProPro 39
         ||||||||||||||||||||||||||
Db  122  TTACTCTCCGGCCCCTCTCTCCCCCTCCA 93


RESULT 88
CK688894
LOCUS       CK688894    733 bp    mRNA    linear   EST 30-MAR-2004
DEFINITION  ZF101-P00047-DBPB-F2_M04 GISZF001_ra Danio rerio cDNA clone
            IMAGE:7151982 5', mRNA sequence.
ACCESSION   CK688894
VERSION     CK688894.1  GI:42441230
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
```

ORGANISM   Danio rerio
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
           Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 733)
AUTHORS   Wei,C., Mathavan,S., Thoreau,H., Lim,L., Lee,C. and Ruan,Y.
TITLE   Genome Institute of Singapore, Zebrafish Gene Collection
JOURNAL   Unpublished (2004)
COMMENT   Contact: Ruan Y
           Cloning and Sequencing
           Genome Institute of Singapore
           60 Biopolis Street, #02-01, Genome, Singapore 138672
           Tel: +65 6478 8073
           Fax: +65 6478 9059
           Email: ruanyj@gis.a-star.edu.sg
           GIS Clone ID: ZF101-P00047-BR2_M04
           PCR Primers
           FORWARD: M13
           BACKWARD: M13
           Plate: ZF101-P00047-BR2 row: M   column: 04
           Seq primer: CCGATAACTTGTATAGCA
           High quality sequence stop: 733.
FEATURES   Location/Qualifiers
   source 1..733
           /organism="Danio rerio"
           /mol_type="mRNA"
           /strain="Singapore local strain"
           /db_xref="taxon:7955"
           /clone="IMAGE:7151982"
           /tissue_type="Embryo"
           /dev_stage="7 Different embryonic stages(From just
           fertilized Embryos to 72 hours just hatched baby fish)"
           /lab_host="DH10B"
           /clone_lib="GIS2F001_ra"
           /note="Vector: pDNR-LIB; Site_1: Sfi A (GGCCATTACGGCC);
           Site_2: Sfi B (GGCCGAGGCGGCC); Priming method: Sfi-(dT)30
           Primed ; Priming sequence:
           5.ATTCTAGAGGCCGAGCGGCGGCCGACATG(T)30VN ; Directionally
           cloned, 5' cloning site: Sfi A site GGCCATTACGGCC ; 5'
           linker/adaptor sequence: 5.AAGCAGTGGTATCAACGCAGAGTCGCC ;
           3' cloning site: Sfi B site GGCCGAGGCGGCC ; 3'
           linker/adaptor sequence: same as the priming sequence ;
           Average insert size: 2kb ; For PCR insert analysis: Use
           M13 Forward and reverse primers ; Library Amplified ;
           Recombinants (inserts): 98% ; Library complexity: 5x106 ;
           Full-length construction (method): SMART, a Clontech
           method The pooled tissue RNA was collected and used to
           construct full length enriched cDNA library and also
           served as template to synthesize complex first strand cDNA
           probe. Two high density colony arrays were made from over
           110K cDNA clones and hybridized with the probes. Low
           intensity clones were selected as they represented rare
           expressed clones .The hybridization intensities for all
           clones span from 0 to 1.8 million counts and the low
           abundant class ranged from 0 to 13,000."
ORIGIN

Alignment Scores:
Pred. No.:   195   Length: 733
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 7   Gaps: 0

US-09-989-890-238 (1-212) x CK688894 (1-733)

Qy   27 SerLeuAlaLeuSerArgProLeuSer 36
           |||||||||||||||||||||||||||
Db   432 AGCCTAGCATTGTTATCAAGACCGTTGAGC 461

RESULT 89
CE265375/c

LOCUS   CE265375   744 bp   DNA   linear   GSS 26-SEP-2003
DEFINITION tigr-gss-dog-1700033935657 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE265375
VERSION   CE265375.1 GI:35986194
KEYWORDS   GSS.
SOURCE   Canis familiaris (dog)
  ORGANISM Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1 (bases 1 to 744)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE   The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT   Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES   Location/Qualifiers
   source   1..744
            /organism="Canis familiaris"
            /mol_type="genomic DNA"
            /strain="Standard Poodle"
            /db_xref="taxon:9615"
            /clone_lib="Dog Library"
            /note="Site 1: BstXI; Libraries were prepared from
            peripheral blood"
ORIGIN

Alignment Scores:
Pred. No.:   197   Length: 744
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 10   Gaps: 0

US-09-989-890-238 (1-212) x CE265375 (1-744)

Qy   23 AlaLeuProSerLeuAlaLeuLeuSer 32
Db   419 GCTCTCCCAAGCAGTCTTGCCCTGCTGAGT 390

RESULT 90
CF386451
LOCUS   CF386451   749 bp   mRNA   linear   EST 29-AUG-2003
DEFINITION RTDR1_14_A09.g1_A015 Loblolly pine roots recovering from drought
            DR1 Pinus_taeda cDNA clone RTDR1_14_A09_A015 5', mRNA sequence.
ACCESSION   CF386451
VERSION   CF386451.1 GI:3434786
KEYWORDS   EST.
SOURCE   Pinus taeda (loblolly pine)
  ORGANISM Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1 (bases 1 to 749)
  AUTHORS   Pratt,L., Cordonnier-Pratt,M.-M., Lorenz,W.W., Dean,J.,
            Gebremedhin,M., Dervinis,C., Martin,T., White,T., Davis,J. and
            Neale,D.
  TITLE   An EST database from loblolly pine (Pinus taeda) roots recovering
            from drought stress
  JOURNAL   Unpublished (2003)
COMMENT   Other_ESTs: RTDR1_14_A09.b1_A015
            Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
RNA prepared and library constructed by W. Walter Lorenz, School of
Forestry, University of Georgia; plant material prepared at the
University of Florida; sequencing done in the Laboratory for
Genomics and Bioinformatics, University of Georgia. Sequence ends
have been trimmed to exclude vector and regions below Phred quality
16. Three-prime sequences are presented as their reverse complement
and have been trimmed to exclude polyA.
Seq primer: JENREV (CAGGAAACAGCTATGACC).

FEATURES   Location/Qualifiers
source   1..749
              /organism="Pinus taeda"
              /mol_type="mRNA"
              /strain="CCLONES"
              /db_xref="taxon:3352"
              /clone="RTDR1_14_A09_A015"
              /lab_host="DH10B-T1 phage-resistant E. coli"
              /clone_lib="Loblolly pine roots recovering from drought
              DR1"
              /note="Vector: pSl1180; Site 1: EcoRI; Site_2: XhoI; The
              library was prepared from polyA+ RNA from loblolly pine
              (Pinus taeda) roots recovering from drought. Water was
              withheld from ramet clones until predawn needle water
              potential reached -1.75 MPa. Plants were well watered on
              day 7 and allowed to recover for 2 days, at which time
              roots were harvested for RNA isolation. Double-stranded
              cDNA was cloned unidirectionally into pSl1180. Inserts
              excised with EcoRI (5' end) and XhoI (3' end)."

ORIGIN

Alignment Scores:
Pred. No.:   199   Length: 749
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 6

US-09-989-890-238 (1-212) x CF386451 (1-749)

Qy   30 LeuLeuSerArgProLeuSerProProPro 39
        ||||||||||||||||||||||||||||||
Db 129 CTCCTTTCAAGACCACTTTCGCCGCCACCT 158

RESULT 91
DR652062
LOCUS   DR652062   749 bp   mRNA   linear   EST 12-JUL-2005
DEFINITION FvN Gibberella moniliformis cDNA clone FVNB408, mRNA
            sequence.
ACCESSION   DR652062
VERSION   DR652062.1 GI:70740538
KEYWORDS   EST.
SOURCE   Gibberella moniliformis
ORGANISM   Gibberella moniliformis
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE   1 (bases 1 to 749)
AUTHORS   Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L,Lee,Y.,
            Utterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D.,
            Kendra,D.F., Town,C.D. and Whitelaw,C.A.
TITLE   Analysis of 87,000 expressed sequence tags reveals alternatively
            spliced introns in multiple genes of the fumonisin gene cluster
JOURNAL   Unpublished (2005)
COMMENT   Contact: Brown, D.W.
            USDA/ARS/NCAUR
            USDA
            1815 N. University St, Peoria, IL 61604, USA
            Tel: 309 681 6230
            Fax: 309 681 6689
            Email: browndw@ncaur.usda.gov
            TIGR sequence name: FVNB408TH
            Seq primer: AAT TAA CCC TCA CTA AAG GG.

FEATURES   Location/Qualifiers
source   1..749
              /organism="Gibberella moniliformis"
              /mol_type="mRNA"
              /strain="m3125"
              /db_xref="taxon:117187"
              /clone="FVNB408"
              /tissue_type="mycelia"
              /clone_lib="FvN"
              /note="Vector: pBlueScript II SK(+) XR; Site 1: EcoRI;
              Site_2: XhoI; anamorph: Fusarium verticillioIdes. Library
              FvN was obtained from RNA derived from a corn meal medium
              culture of strain M-3125. These cultures were prepared by
              inoculating an autoclaved mixture of 25 g corn meal and 5
              ml distilled with 5 ml of water containing 5 x 10e7
              condia. The inoculated medium was mixed thoroughly,
              distributed equally into two 100-mm plastic petri dishes,
              and incubated at room temperature for four or six days."

ORIGIN

Alignment Scores:
Pred. No.:   199   Length: 749
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 8

US-09-989-890-238 (1-212) x DR652062 (1-749)

Qy   30 LeuLeuSerArgProLeuSerProProPro 39
        ||||||||||||||||||||||||||||||
Db 687 CTCCTTTCAAGACCACTTTCGCCGCCACCT 716

RESULT 92
DR647955
LOCUS   DR647955   757 bp   mRNA   linear   EST 12-JUL-2005
DEFINITION EST1038072 FvN Gibberella moniliformis cDNA clone FVNAD49, mRNA
            sequence.
ACCESSION   DR647955
VERSION   DR647955.1 GI:70736430
KEYWORDS   EST.
SOURCE   Gibberella moniliformis
ORGANISM   Gibberella moniliformis
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE   1 (bases 1 to 757)
AUTHORS   Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L, Lee,Y.,
            Utterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D.,
            Kendra,D.F., Town,C.D. and Whitelaw,C.A.
TITLE   Analysis of 87,000 expressed sequence tags reveals alternatively
            spliced introns in multiple genes of the fumonisin gene cluster
JOURNAL   Unpublished (2005)
COMMENT   Contact: Brown, D.W.
            USDA/ARS/NCAUR
            USDA
            1815 N. University St, Peoria, IL 61604, USA
            Tel: 309 681 6230
            Fax: 309 681 6689
            Email: browndw@ncaur.usda.gov
            TIGR sequence name: FVNAD49TH
            Seq primer: AAT TAA CCC TCA CTA AAG GG.

FEATURES   Location/Qualifiers
source   1..757
              /organism="Gibberella moniliformis"
              /mol_type="mRNA"
              /strain="m3125"
              /db_xref="taxon:117187"
              /clone="FVNAD49"
              /tissue_type="mycelia"

```
/clone_lib="FvN"
/note="Vector: pBlueScript II SK(+) XR; Site 1: EcoRI;
Site 2: XhoI; anamorph: Fusarium verticillioides. Library
FvN was obtained from RNA derived from a corn meal medium
culture of strain M-3125. These cultures were prepared by
inoculating an autoclaved mixture of 25 g corn meal and 5
ml distilled water with 5 ml of water containing 5 x 10e7
condia. The inoculated medium was mixed thoroughly,
distributed equally into two 100-mm plastic petri dishes,
and incubated at room temperature for four or six days."

ORIGIN

Alignment Scores:
Pred. No.:                201          Length:           757
Score:                  10.00          Matches:           10
Percent Similarity:    100.0%          Conservative:       0
Best Local Similarity: 100.0%          Mismatches:         0
Query Match:             4.7%          Indels:             0
DB:                         8          Gaps:               0

US-09-989-890-238 (1-212) x DR647955 (1-757)

Qy    30 LeuLeuSerArgProLeuSerProProPro 39
            ||||||||||||||||||||||||||||
Db   640 CTCCTTTCAAGACCACTTTCGCCGCCACCT 669

RESULT 93
CX773156
LOCUS        CX773156            758 bp    mRNA       linear    EST 25-JAN-2005
DEFINITION   UI-EH-HG0-aak-o-08-0-UI.s1 UI-EH-HG0 Emiliania huxleyi cDNA clone
             UI-EH-HG0-aak-o-08-0-UI 3', mRNA sequence.
ACCESSION    CX773156
VERSION      CX773156.1  GI:58183509
KEYWORDS     EST.
SOURCE       Emiliania huxleyi
  ORGANISM   Emiliania huxleyi
             Eukaryota; Haptophyceae; Isochrysidales; Emiliania.
REFERENCE    1  (bases 1 to 758)
  AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE      Normalization and subtraction: two approaches to facilitate gene
             discovery
  JOURNAL    Genome Res. 6 (9), 791-806 (1996)
   PUBMED    8889548
COMMENT      Contact: Soares, MB
             Coordinated Laboratory for Computational Genomics
             University of Iowa
             375 Newton Road, 4156  MERBF, Iowa City, IA 52242, USA
             Tel: 319 335 8250
             Fax: 319 335 9565
             Email: bento-soares@uiowa.edu
             Tissue Procurement: Provasoli-Guillard National Center for Culture
             of Marine Phytoplankton (CCMP)
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
             http://genome.uiowa.edu/distribution/dinoflagellate.html
             The following repetitive elements were found in this cDNA
             sequence: 708-758, >GC_rich#Low_complexity
             Seq primer: M13 FORWARD
             POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..758
                     /organism="Emiliania huxleyi"
                     /mol_type="mRNA"
                     /strain="CCMP 371"
                     /db_xref="taxon:2903"
                     /clone="UI-EH-HG0-aak-o-08-0-UI"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-EH-HG0"
                     /note="Vector: pT7T3-Pac (Pharmacia) with a modified
                     polylinker; Site 1: EcoR I; Site 2: Not I; The library was
                     constructed according to Bonaldo, Lennon and Soares,
```

```
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
GAAGGCTAGT. Tissue was obtained from the
Provasoli-Guillard National Center for Culture of Marine
Phytoplankton (CCMP)
TAG_TISSUB=Emiliania huxleyi coccolithophorid
TAG_LIB=UI-EH-HG0
TAG_SEQ=GAAGGCTAGT"

ORIGIN

Alignment Scores:
Pred. No.:                201          Length:           758
Score:                  10.00          Matches:           10
Percent Similarity:    100.0%          Conservative:       0
Best Local Similarity: 100.0%          Mismatches:         0
Query Match:             4.7%          Indels:             0
DB:                         8          Gaps:               0

US-09-989-890-238 (1-212) x CX773156 (1-758)

Qy    32 SerArgProLeuSerProProProAlaAla 41
            |||||||||||||||||||||||||||||
Db    93 AGCAGGCCGCTCTCACCGCCCCGCCGGGCG 122

RESULT 94
DR646867
LOCUS        DR646867
DEFINITION   EST1036984 FvN Gibberella moniliformis cDNA clone FVNA630, mRNA
             sequence.                           760 bp    mRNA       linear    EST 12-JUL-2005
ACCESSION    DR646867
VERSION      DR646867.1  GI:70735342
KEYWORDS     EST.
SOURCE       Gibberella moniliformis
  ORGANISM   Gibberella moniliformis
             Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
             Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE    1  (bases 1 to 760)
  AUTHORS    Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L,Lee,Y.,
             Utterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D.,
             Kendra,D.F., Town,C.D. and Whitelaw,C.A.
  TITLE      Analysis of 87,000 expressed sequence tags reveals alternatively
             spliced introns in multiple genes of the fumonisin gene cluster
  JOURNAL    Unpublished (2005)
COMMENT      Contact: Brown, D.W.
             USDA/ARS/NCAUR
             USDA
             1815 N. University St, Peoria, IL 61604, USA
             Tel: 309 681 6230
             Fax: 309 681 6689
             Email: browndw@ncaur.usda.gov
             TIGR sequence name: FVNA630TH
             Seq primer: AAT TAA CCC TCA CTA AAG GG.
FEATURES             Location/Qualifiers
     source          1..760
                     /organism="Gibberella moniliformis"
                     /mol_type="mRNA"
                     /strain="m3125"
                     /db_xref="taxon:117187"
                     /clone="FVNA630"
                     /tissue_type="mycelia"
                     /clone_lib="FvN"
                     /note="Vector: pBlueScript II SK(+) XR; Site 1: EcoRI;
                     Site 2: XhoI; anamorph: Fusarium verticillioides. Library
                     FvN was obtained from a corn meal medium
                     culture of strain M-3125. These cultures were prepared by
                     inoculating an autoclaved mixture of 25 g corn meal and 5
                     ml distilled water with 5 ml of water containing 5 x 10e7
```

ORIGIN   condia. The inoculated medium was mixed thoroughly,
               distributed equally into two 100-mm plastic petri dishes,
               and incubated at room temperature for four or six days."

Alignment Scores:
Pred. No.:   201   Length: 760
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x DR646867 (1-760)

Qy   30 LeuLeuSerArgProLeuSerProPro 39
       |||||||||||||||||||||||||||
Db   52 CTCCTTTCAAGACCACTTTCGCCGCCACCT 81

RESULT 95
CK473171
LOCUS   CK473171   769 bp   mRNA   linear   EST 14-JAN-2004
DEFINITION AGENCOURT 17617923 NIH_MGC_236 Rattus norvegicus cDNA clone
          IMAGE:7128510 5', mRNA sequence.
ACCESSION CK473171
VERSION   CK473171.1 GI:40817269
KEYWORDS   EST.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 769)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished (1999)
JOURNAL   Contact: Daniela S. Gerhard, Ph.D.
COMMENT   Office of Cancer Genomics
          National Cancer Institute / NIH
          Bldg. 31 Rm10A07 Bethesda, MD 20892
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Howard Jacobs
          cDNA Library Preparation: Express Genomics
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM15024 row: k column: 04
          High quality sequence stop: 491.
FEATURES   Location/Qualifiers
 source   1..769
              /organism="Rattus norvegicus"
              /mol_type="mRNA"
              /db_xref="taxon:10116"
              /clone="IMAGE:7128510"
              /tissue_type="kidney, pooled"
              /lab_host="DH10B TonA"
              /clone_lib="NIH_MGC_236"
              /note="Organ: kidney; Vector: pExpress-1; Site 1: EcoRV;
              Site 2: NotI; RNA obtained from pooled kidney tissue from
              a mix of male and female animals at 8 wk old. Tissues were
              snap-frozen and kept at -80C for two days before RNA
              extraction and purification (Tri-reagent method). cDNA was
              primed using oligo-dT primer:
              5'-pGACTAGTTCTAGATCGGACGCGAGCGGCCGCCC(T)25-3' and cloned into
              the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb
              resulted in an average insert size of 2.2 kb. This primary
              library is normalized (non-normalized primary library is
              NIH_MGC 235) and was constructed by Express Genomics
              (Frederick, MD). Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:
Pred. No.:   207   Length: 783

Alignment Scores:
Pred. No.:   204   Length: 769
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 7   Gaps: 0

US-09-989-890-238 (1-212) x CK473171 (1-769)

Qy   60 AlaAlaAlaGlyIleAlaSerSerAlaValGlu 69
       ||||||||||||||||||||||||||||||||
Db   352 GCTGCAGGGATTGCCTCCAGCGCTGTGGAG 381

RESULT 96
BQ180749/c
LOCUS   BQ180749   783 bp   mRNA   linear   EST 30-APR-2002
DEFINITION UI-M-EX0-bxc-o-01-0-UI.r1 NIH_BMAP_EX0 Mus musculus cDNA clone
          IMAGE:5706456 5', mRNA sequence.
ACCESSION BQ180749
VERSION   BQ180749.1 GI:20356241
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 783)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Dr. James Lin, Univeristy of Iowa
          cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
          cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
          DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
          Clone Distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          This clone was contributed by the Brain Molecular Anatomy Project
          (BMAP)
          Seq primer: pYX-5.
FEATURES   Location/Qualifiers
 source   1..783
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="C57Bl/6"
              /db_xref="taxon:10090"
              /clone="IMAGE:5706456"
              /tissue type="whole brain"
              /dev_stage="embryo 15.5 dpc"
              /lab_host="DH10B (T1 phage resistant)"
              /clone_lib="NIH_BMAP_EX0"
              /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
              Site_2: Not I; The library was constructed according to
              Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
              1996. Denatured mRNA was size fractionated on a 1% agarose
              gel. First strand cDNA synthesis was primed with an
              oligo-dT primer containing a Not I site. Double stranded
              cDNA was size selected according to mRNA size fraction,
              ligated with EcoR I adaptor, digested with Not I, and then
              cloned directionally into pYX-Asc vector. The library tag
              sequence located between the Not I site and the polyA
              tail, is GTGCGTGGAA. This library was created for the
              University of Iowa Mouse Brain Molecular Anatomy Project
              (BMAP): 'Gene Discovery in the Developing Mouse Nervous
              System', supported by National Instututes of Mental Health
              (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

```
Score:                  10.00        Matches:          10
Percent Similarity:     100.0%       Conservative:      0
Best Local Similarity:  100.0%       Mismatches:        0
Query Match:            4.7%         Indels:            0
DB:                     3            Gaps:              0

US-09-989-890-238 (1-212) x BQ180749 (1-783)

Qy   52  AlaGlyLeuProSerAlaSerAlaAlaAla  61
         ||||||||||||||||||||||||||||||
Db   38  GCTGGGCTGCCTAGTGCCAGCGCTGCG  9

RESULT 97
CK478920
LOCUS       CK478920                783 bp    mRNA     linear   EST 14-JAN-2004
DEFINITION  AGENCOURT_17582913 NIH_MGC 232 Rattus norvegicus cDNA clone
            IMAGE:7121341 5', mRNA sequence.
ACCESSION   CK478920
VERSION     CK478920.1  GI:40823054
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 783)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Howard Jacobs
            cDNA Library Preparation: Express Genomics
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM15005 row: p column: 11
            High quality sequence stop: 736.
FEATURES             Location/Qualifiers
     source          1..783
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7121341"
                     /tissue_type="lung, pooled"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC 232"
                     /note="Organ: lung; Vector: pExpress-1; Site_1: EcoRV;
                     Site_2: NotI; RNA obtained from pooled lung tissue from a
                     mix of male and female animals at 8 wk old. Tissues were
                     snap-frozen and kept at -80C for two days before RNA
                     extraction and purification (Tri-reagent method). cDNA was
                     primed using oligo-dT primer;
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb
                     resulted in an average insert size of 2.3 kb. This primary
                     library is normalized (non-normalized primary library is
                     NIH_MGC 231) and was constructed by Express Genomics
                     (Frederick, MD). Note: this is a NIH_MGC library."
ORIGIN

Alignment Scores:
Pred. No.:              207          Length:          783
Score:                  10.00        Matches:          10
Percent Similarity:     100.0%       Conservative:      0
Best Local Similarity:  100.0%       Mismatches:        0
Query Match:            4.7%         Indels:            0
DB:                     7            Gaps:              0

US-09-989-890-238 (1-212) x CK478920 (1-783)

Qy   60  AlaAlaGlyIleAlaSerSerAlaValGlu  69
         ||||||||||||||||||||||||||||||
Db  371  GCTGCAGGGATTGCCTCCAGCGCTGTGGAG  400

RESULT 98
DR656203/c
LOCUS       DR656203                785 bp    mRNA     linear   EST 12-JUL-2005
DEFINITION  EST1046320 FvN Gibberella moniliformis cDNA clone FVNBU54, mRNA
            sequence.
ACCESSION   DR656203
VERSION     DR656203.1  GI:70744679
KEYWORDS    EST.
SOURCE      Gibberella moniliformis
ORGANISM    Gibberella moniliformis
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE   1  (bases 1 to 785)
AUTHORS     Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L, Lee,Y.,
            Utterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D.,
            Kendra,D.F., Town,C.D. and Whitelaw,C.A.
TITLE       Analysis of 87,000 expressed sequence tags reveals alternatively
            spliced introns in multiple genes of the fumonisin gene cluster
JOURNAL     Unpublished (2005)
COMMENT     Contact: Brown, D.W.
            USDA/ARS/NCAUR
            USDA
            1815 N. University St, Peoria, IL 61604, USA
            Tel: 309 681 6230
            Fax: 309 681 6689
            Email: browndw@ncaur.usda.gov
            TIGR sequence name: FVNBU54TV
            Seq primer: GTA ATA CGA CTC ACT ATA GGG C.
FEATURES             Location/Qualifiers
     source          1..785
                     /organism="Gibberella moniliformis"
                     /mol_type="mRNA"
                     /strain="m3125"
                     /db_xref="taxon:117187"
                     /clone="FVNBU54"
                     /tissue_type="mycelia"
                     /clone_lib="FvN"
                     /note="Vector: pBlueScript II SK(+) XR; Site_1: EcoRI;
                     Site_2: XhoI; anamorph: Fusarium verticillioides. Library
                     FvN was obtained from RNA derived from a corn meal medium
                     culture of strain M-3125. These cultures were prepared by
                     inoculating an autoclaved mixture of 25 g corn meal and 5
                     ml distilled water with 5 ml of water containing 5 x 10e7
                     condia. The inoculated medium was mixed thoroughly,
                     distributed equally into two 100-mm plastic petri dishes,
                     and incubated at room temperature for four or six days."
ORIGIN

Alignment Scores:
Pred. No.:              208          Length:          785
Score:                  10.00        Matches:          10
Percent Similarity:     100.0%       Conservative:      0
Best Local Similarity:  100.0%       Mismatches:        0
Query Match:            4.7%         Indels:            0
DB:                     8            Gaps:              0

US-09-989-890-238 (1-212) x DR656203 (1-785)

Qy   30  LeuLeuSerArgProLeuSerProProPro  39
         ||||||||||||||||||||||||||||||
Db  717  CTCCTTTCAAGACCACTTTCGCGCCACCT  688

RESULT 99
BZ812231
LOCUS       BZ812231                809 bp    DNA      linear   GSS 17-MAR-2003
DEFINITION  PUFCY39TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa291G05,
```

genomic survey sequence.
ACCESSION   BZ812231
VERSION   BZ812231.1 GI:29024584
KEYWORDS   GSS.
SOURCE   Zea mays
ORGANISM   Zea mays
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
             clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 809)
AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
             Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
             Bennetzen,J.
TITLE   Maize Genomics Consortium
JOURNAL   Unpublished (2003)
COMMENT   Other GSSs: PUFCY39TB
             Contact: Cathy Whitelaw
             TIGR
             9712 Medical Center Drive, Rockville, MD 20850, USA
             Tel: 301-838-5843
             Fax: 301-838-0208
             Email: whitelaw@tigr.org
             Seq primer: TF
             Class: sheared ends.
FEATURES   Location/Qualifiers
     source 1..809
             /organism="Zea mays"
             /mol_type="genomic DNA"
             /strain="B73"
             /db_xref="taxon:4577"
             /clone="ZMMBTa291G05"
             /clone_lib="ZM_0.6-1.0_KB"
             /note="vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
             CoT selected genomic DNA library"

ORIGIN

Alignment Scores:
Pred. No.:   214   Length: 809
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 9   Gaps: 0

US-09-989-890-238 (1-212) x BZ812231 (1-809)

Qy 50 SerGlyAlaGlyLeuProSerAlaSerAla 59
        ||||||||||||||||||||||||||||||
Db 315 TCTGGAGCTGGGCTTCCTAGTGCGTCGGCC 344

RESULT 100
BQ770347/c
LOCUS   BQ770347   810 bp   mRNA   linear   EST 26-JUL-2002
DEFINITION   UI-M-FI0-byu-j-06-0-UI.r1 NIH_BMAP_FI0 Mus musculus cDNA clone
             IMAGE:5702501 5', mRNA sequence.
ACCESSION   BQ770347
VERSION   BQ770347.1 GI:21978821
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 810)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

COMMENT   Clone Distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
             (BMAP)
             Seq primer: pYX-5.
FEATURES   Location/Qualifiers
     source 1..810
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6"
             /db_xref="taxon:10090"
             /clone="IMAGE:5702501"
             /tissue_type="whole brain"
             /dev_stage="embryo 12.5dpc"
             /lab_host="DH10B (T1 phage resistant)"
             /clone_lib="NIH_BMAP_FI0"
             /note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
             Site 2: Not I; The library was constructed according
             Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
             1996. Denatured RNA was size fractionated on a 1% agarose
             gel.First strand cDNA synthesis was primed with oligo-dT
             primer containing a Not I site. Double strand cDNA was
             size selected according to mRNA size fraction, ligated
             with EcoR I adaptor, digested with NotI and then cloned
             directionally into pYX-Asc vector. The library tag
             sequence located between the Not I site and the polyA tail
             is CAGCCACGAC. This library was created for the University
             Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
             Developing Mouse Nervous System', supported by National
             Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
             program coordinator."

ORIGIN

Alignment Scores:
Pred. No.:   214   Length: 810
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: 5   Gaps: 0

US-09-989-890-238 (1-212) x BQ770347 (1-810)

Qy 52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
        ||||||||||||||||||||||||||||||
Db 32 GCTGGGCTGCCTAGTGCCAGCCGCTGCTGCG 3

Search completed: March 17, 2006, 08:19:15
Job time : 3913 secs

THIS PAGE LEFT BLANK

```
                GenCore version 5.1.7
         Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:    March 17, 2006, 06:24:31 ; Search time 505 Seconds
                                       (without alignments)
                                       746.224 Million cell updates/sec

Title:         US-09-989-890-238
Perfect score: 212
Sequence:      1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table: OLIGO
               Xgapop 60.0 , Xgapext 60.0
               Ygapop 60.0 , Ygapext 60.0
               Fgapop  6.0 , Fgapext  7.0
               Delop   6.0 , Delext   7.0

Searched:   1303057 seqs, 888780828 residues
Word size:  1
Total number of hits satisfying chosen parameters:        2598827

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=x1h
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095518_17602/app_query.fasta_1
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=oligo_p2n.rni -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=500 -DOCALIGN=200 -THR_SCORE=quality-THR_MIN=1
-ALIGN=100 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abss05h
-USER=US09989890 @CGN_1_1_193 @runat_16032006_095518_17602 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
   1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
   2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
   3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
   4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
   5: /cgn2_6/ptodata/1/ina/6H_COMB.seq:*
   6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
   7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
   8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
   9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	115	54.2	1785	3	US-10-104-047-799	Sequence 799, App
2	11	5.2	194937	3	US-09-949-016-17032	Sequence 17032, A
3	11	5.2	194937	3	US-09-949-016-17033	Sequence 17033, A
4	10	4.7	31861	3	US-09-949-016-12803	Sequence 12803, A
5	10	4.7	31861	3	US-09-949-016-13967	Sequence 13967, A
c 6	9	4.2	601	3	US-09-949-016-30920	Sequence 30920, A
c 7	9	4.2	601	3	US-09-949-016-30921	Sequence 30921, A
c 8	9	4.2	601	3	US-09-949-016-30922	Sequence 30922, A

	No.						ID	Description
c	9	c	9	4.2	8	601	US-09-949-016-30923	Sequence 30923, A
c	10	c	9	4.2	8	601	US-09-949-016-30924	Sequence 30924, A
c	11	c	9	4.2	8	601	US-09-949-016-168951	Sequence 168951,
c	12	c	9	4.2	8	601	US-09-949-016-168952	Sequence 168952,
c	13	c	9	4.2	8	601	US-09-949-016-168953	Sequence 168953,
c	14	c	9	4.2	8	601	US-09-949-016-168954	Sequence 168954,
c	15	c	9	4.2	8	601	US-09-949-016-168955	Sequence 168955,
c	16	c	9	4.2	8	1377	US-09-902-540-8319	Sequence 8319, Ap
c	17	c	9	4.2	8	1716	US-09-902-540-9034	Sequence 9034, Ap
c	18	c	9	4.2	8	7874	US-09-902-540-868	Sequence 868, App
c	19	c	9	4.2	8	9198	US-09-902-540-966	Sequence 966, App
c	20	c	9	4.2	8	9825	US-09-949-016-15117	Sequence 15117, A
c	21	c	9	4.2	8	10432	US-09-919-172-97	Sequence 97, Appl
c	22	c	9	4.2	8	10432	US-09-976-594-21	Sequence 21, Appl
c	23	c	9	4.2	8	10432	US-09-919-039-20	Sequence 20, Appl
c	24	c	9	4.2	8	16013	US-09-949-016-12988	Sequence 12988, A
c	25	c	9	4.2	8	39498	US-09-949-016-12410	Sequence 12410, A
c	26	c	9	4.2	8	39498	US-09-949-016-16505	Sequence 16505, A
c	27	c	9	4.2	8	51252	US-09-949-016-16348	Sequence 16348, A
c	28	c	9	4.2	8	79595	US-09-949-016-15318	Sequence 15318, A
c	29	c	9	4.2	8	79634	US-09-949-016-15642	Sequence 15642, A
c	30	c	9	4.2	8	86414	US-09-949-016-12345	Sequence 12345, A
c	31	c	9	4.2	8	86414	US-09-949-016-15758	Sequence 15758, A
c	32	c	9	4.2	8	148794	US-09-949-016-12751	Sequence 12751, A
c	33	c	9	4.2	8	455726	US-09-949-016-14157	Sequence 14157, A
c	34	c	9	4.2	8	481115	US-09-949-016-11940	Sequence 11940, A
	35	c	8	3.8	8	50	US-10-131-827-69	Sequence 69, Appl
	36	c	8	3.8	8	50	US-10-131-827-4540	Sequence 4540, Ap
	37	c	8	3.8	8	55	US-08-474-633A-96	Sequence 96, Appl
	38	c	8	3.8	8	55	US-08-474-633A-97	Sequence 97, Appl
	39	c	8	3.8	8	55	US-08-737-524B-20	Sequence 20, Appl
	40	c	8	3.8	8	55	US-08-737-524B-21	Sequence 21, Appl
	41	c	8	3.8	8	55	US-08-823-771-96	Sequence 96, Appl
	42	c	8	3.8	8	55	US-08-823-771-97	Sequence 97, Appl
	43	c	8	3.8	6	55	PCT-US93-02480-19	Sequence 19, Appl
	44	c	8	3.8	6	55	PCT-US93-02480-20	Sequence 20, Appl
	45	c	8	3.8	6	55	PCT-US95-08501-5	Sequence 5, Aapli
	46	c	8	3.8	6	55	PCT-US95-08501-6	Sequence 6, Aapli
	47	c	8	3.8	2	180	US-09-513-909C-12253	Sequence 12253, A
	48	c	8	3.8	3	270	US-09-949-016-807	Sequence 807, App
	49	c	8	3.8	6	336	US-09-949-016-2217	Sequence 2217, Ap
	50	c	8	3.8	3	441	US-09-123-912-97	Sequence 97, Appl
	51	c	8	3.8	3	441	US-09-643-597-97	Sequence 97, Appl
	52	c	8	3.8	3	441	US-09-480-884A-97	Sequence 97, Appl
	53	c	8	3.8	3	441	US-09-542-615A-97	Sequence 97, Appl
	54	c	8	3.8	3	441	US-09-606-421B-97	Sequence 97, Appl
	55	c	8	3.8	3	441	US-09-221-107-97	Sequence 97, Appl
	56	c	8	3.8	3	441	US-09-466-396A-97	Sequence 97, Appl
	57	c	8	3.8	3	441	US-09-476-496A-97	Sequence 97, Appl
	58	c	8	3.8	3	441	US-09-630-940B-97	Sequence 97, Appl
	59	c	8	3.8	3	441	US-09-285-479-97	Sequence 97, Appl
	60	c	8	3.8	3	441	US-10-007-700-97	Sequence 97, Appl
	61	c	8	3.8	3	479	US-09-270-767-7252	Sequence 7252, Ap
	62	c	8	3.8	3	479	US-09-270-767-21534	Sequence 21534, A
c	63	c	8	3.8	3	528	US-08-913-362-29	Sequence 29, Appl
c	64	c	8	3.8	3	569	US-09-542-615A-97	Sequence 191, App
	65	c	8	3.8	3	578	US-08-991-789A-222	Sequence 222, App
	66	c	8	3.8	3	578	US-09-062-451-222	Sequence 222, App
	67	c	8	3.8	3	578	US-09-598-326-222	Sequence 222, App
	68	c	8	3.8	3	578	US-09-289-198-222	Sequence 222, App
	69	c	8	3.8	3	578	US-09-429-755-222	Sequence 222, App
	70	c	8	3.8	3	578	US-09-295-295-222	Sequence 222, App
	71	c	8	3.8	3	578	US-09-534-025A-222	Sequence 222, App
	72	c	8	3.8	3	600	US-09-123-912-98	Sequence 98, Appl
	73	c	8	3.8	3	600	US-09-643-597-98	Sequence 98, Appl
	74	c	8	3.8	3	600	US-09-480-884A-98	Sequence 98, Appl
	75	c	8	3.8	3	600	US-09-542-615A-98	Sequence 98, Appl
	76	c	8	3.8	3	600	US-09-606-421B-98	Sequence 98, Appl
	77	c	8	3.8	3	600	US-09-221-107-98	Sequence 98, Appl
	78	c	8	3.8	3	600	US-09-466-396A-98	Sequence 98, Appl
	79	c	8	3.8	3	600	US-09-476-496A-98	Sequence 98, Appl
	80	c	8	3.8	3	600	US-09-630-940B-98	Sequence 98, Appl
	81	c	8	3.8	3	600	US-09-285-479-98	Sequence 98, Appl

	#					Patent		Sequence
c	82	8	3.8	600	3	US-10-007-700-98		Sequence 98, Appl
c	83	8	3.8	601	3	US-09-949-016-25231		Sequence 25231, A
	84	8	3.8	601	3	US-09-949-016-30352		Sequence 30352, A
	85	8	3.8	601	3	US-09-949-016-33707		Sequence 33707, A
	86	8	3.8	601	3	US-09-949-016-38879		Sequence 38879, A
	87	8	3.8	601	3	US-09-949-016-72530		Sequence 72530, A
	88	8	3.8	601	3	US-09-949-016-76311		Sequence 76311, A
	89	8	3.8	601	3	US-09-949-016-88437		Sequence 88437, A
	90	8	3.8	601	3	US-09-949-016-92611		Sequence 92611, A
	91	8	3.8	601	3	US-09-949-016-92612		Sequence 92612, A
	92	8	3.8	601	3	US-09-949-016-92613		Sequence 92613, A
	93	8	3.8	601	3	US-09-949-016-123793		Sequence 123793,
c	94	8	3.8	601	3	US-09-949-016-133788		Sequence 133788,
c	95	8	3.8	601	3	US-09-949-016-133789		Sequence 133789,
	96	8	3.8	601	3	US-09-949-016-141154		Sequence 141154,
c	97	8	3.8	601	3	US-09-949-016-141155		Sequence 141155,
	98	8	3.8	601	3	US-09-949-016-147427		Sequence 147427,
c	99	8	3.8	601	3	US-09-949-016-147591		Sequence 147591,
	100	8	3.8	601	3	US-09-949-016-157906		Sequence 157906,
	101	8	3.8	601	3	US-09-949-016-157907		Sequence 157907,
	102	8	3.8	601	3	US-09-949-016-157908		Sequence 157908,
	103	8	3.8	601	3	US-09-949-016-170627		Sequence 170627,
	104	8	3.8	601	3	US-09-949-016-179321		Sequence 179321,
	105	8	3.8	601	3	US-09-949-016-183714		Sequence 183714,
	106	8	3.8	601	3	US-09-949-016-190102		Sequence 190102,
	107	8	3.8	601	3	US-09-949-016-190103		Sequence 190103,
c	108	8	3.8	660	3	US-09-949-002-33		Sequence 33, Appl
c	109	8	3.8	708	3	US-09-252-991A-3374		Sequence 3374, Ap
	110	8	3.8	710	3	US-08-913-362-3		Sequence 3, Appli
c	111	8	3.8	810	3	US-08-913-362-7		Sequence 7, Appli
c	112	8	3.8	830	3	US-08-913-362-1		Sequence 1, Appli
	113	8	3.8	840	3	US-09-252-991A-14465		Sequence 14465, A
	114	8	3.8	842	3	US-09-270-767-11918		Sequence 11918, A
	115	8	3.8	850	3	US-08-913-362-5		Sequence 5, Appli
	116	8	3.8	934	2	US-08-474-379C-81		Sequence 81, Appl
	117	8	3.8	934	2	US-09-146-249A-81		Sequence 81, Appl
	118	8	3.8	936	2	US-08-206-188B-81		Sequence 81, Appl
	119	8	3.8	936	3	US-09-270-767-4464		Sequence 4464, Ap
	120	8	3.8	936	3	US-09-270-767-19746		Sequence 19746, A
	121	8	3.8	939	3	US-09-252-991A-16426		Sequence 16426, A
c	122	8	3.8	1026	3	US-09-949-002-33		Sequence 33, Appl
c	123	8	3.8	1049	3	US-09-069-023-13		Sequence 13, Appli
c	124	8	3.8	1065	3	US-09-252-991A-16320		Sequence 16320, A
	125	8	3.8	1065	3	US-09-533-559-7080		Sequence 7080, Ap
c	126	8	3.8	1146	2	US-08-227-800A-1		Sequence 1, Appli
c	127	8	3.8	1146	2	US-08-921-954-1		Sequence 1, Appli
c	128	8	3.8	1149	2	US-08-474-177-3		Sequence 3, Appli
c	129	8	3.8	1149	2	US-08-487-033-3		Sequence 3, Appli
c	130	8	3.8	1149	2	US-08-480-810-3		Sequence 3, Appli
c	131	8	3.8	1149	2	US-08-508-735-3		Sequence 3, Appli
c	132	8	3.8	1149	2	US-08-848-251-3		Sequence 3, Appli
c	133	8	3.8	1149	2	US-08-486-047-3		Sequence 3, Appli
c	134	8	3.8	1149	2	US-09-120-130-3		Sequence 3, Appli
c	135	8	3.8	1149	2	US-09-115-252-3		Sequence 3, Appli
c	136	8	3.8	1149	2	US-08-986-515-3		Sequence 3, Appli
c	137	8	3.8	1149	2	US-08-120-128-3		Sequence 3, Appli
	138	8	3.8	1149	2	US-09-120-129-3		Sequence 3, Appli
c	139	8	3.8	1149	2	US-09-201-139-3		Sequence 3, Appli
	140	8	3.8	1149	2	US-09-120-131-3		Sequence 3, Appli
c	141	8	3.8	1155	2	US-07-688-352C-45		Sequence 45, Appl
c	142	8	3.8	1155	2	US-08-474-379C-45		Sequence 45, Appl
c	143	8	3.8	1155	2	US-08-146-249A-45		Sequence 45, Appl
	144	8	3.8	1155	2	US-08-206-188B-45		Sequence 45, Appl
	145	8	3.8	1155	6	PCT-US91-02714-42		Sequence 42, Appl
	146	8	3.8	1155	6	PCT-US91-02714-53		Sequence 53, Appl
c	147	8	3.8	1209	3	US-10-166-653-17		Sequence 17, Appl
c	148	8	3.8	1221	3	US-09-252-991A-9339		Sequence 9339, Ap
c	149	8	3.8	1298	3	US-09-311-784A-25		Sequence 25, Appl
	150	8	3.8	1320	3	US-09-328-352-3080		Sequence 3080, Ap
c	151	8	3.8	1332	3	US-09-252-991A-15961		Sequence 15961, A
	152	8	3.8	1341	3	US-09-252-991A-16544		Sequence 16544, A
c	153	8	3.8	1347	3	US-09-489-039A-2226		Sequence 2226, Ap
c	154	8	3.8	1365	3	US-09-252-991A-9407		Sequence 9407, Ap

	#					Patent	Sequence
c	155	8	3.8	1368	3	US-09-902-540-7805	Sequence 7805, Ap
	156	8	3.8	1377	2	US-08-112-817C-1	Sequence 1, Appli
c	157	8	3.8	1386	3	US-09-433-248A-1	Sequence 1, Appli
c	158	8	3.8	1420	3	US-09-589-981-1	Sequence 1, Appli
c	159	8	3.8	1420	3	US-09-457-568-3	Sequence 3, Appli
c	160	8	3.8	1420	3	US-09-457-646-3	Sequence 3, Appli
c	161	8	3.8	1420	3	US-09-902-572-1	Sequence 1, Appli
c	162	8	3.8	1420	3	US-09-516-065-3	Sequence 3, Appli
c	163	8	3.8	1449	3	US-09-252-991A-15134	Sequence 15134, A
c	164	8	3.8	1459	3	US-09-569-611C-4	Sequence 4, Appli
c	165	8	3.8	1470	2	US-08-124-674-1	Sequence 1, Appli
c	166	8	3.8	1470	2	US-08-589-893-1	Sequence 1, Appli
c	167	8	3.8	1470	2	US-08-589-893-3	Sequence 3, Appli
c	168	8	3.8	1470	2	US-08-589-893-5	Sequence 5, Appli
c	169	8	3.8	1470	2	US-08-589-893-7	Sequence 7, Appli
c	170	8	3.8	1470	2	US-08-589-893-9	Sequence 9, Appli
c	171	8	3.8	1470	2	US-08-589-893-11	Sequence 11, Appl
c	172	8	3.8	1470	2	US-08-589-893-13	Sequence 13, Appl
c	173	8	3.8	1470	2	US-08-589-893-15	Sequence 15, Appl
c	174	8	3.8	1470	2	US-08-589-893-17	Sequence 17, Appl
c	175	8	3.8	1470	2	US-08-589-893-19	Sequence 19, Appl
c	176	8	3.8	1470	2	US-08-589-893-21	Sequence 21, Appl
c	177	8	3.8	1470	2	US-08-589-893-23	Sequence 23, Appl
c	178	8	3.8	1470	2	US-09-020-991-1	Sequence 1, Appli
c	179	8	3.8	1470	2	US-09-020-991-3	Sequence 3, Appli
c	180	8	3.8	1470	2	US-09-020-991-5	Sequence 5, Appli
c	181	8	3.8	1470	2	US-09-020-991-7	Sequence 7, Appli
c	182	8	3.8	1470	2	US-09-020-991-9	Sequence 9, Appli
c	183	8	3.8	1470	2	US-09-020-991-11	Sequence 11, Appl
c	184	8	3.8	1470	2	US-09-020-991-13	Sequence 13, Appl
c	185	8	3.8	1470	2	US-09-020-991-15	Sequence 15, Appl
c	186	8	3.8	1470	2	US-09-020-991-17	Sequence 17, Appl
c	187	8	3.8	1470	2	US-09-020-991-19	Sequence 19, Appl
c	188	8	3.8	1470	2	US-09-020-991-21	Sequence 21, Appl
c	189	8	3.8	1470	2	US-09-020-991-23	Sequence 23, Appl
c	190	8	3.8	1470	2	US-09-062-890-1	Sequence 1, Appli
c	191	8	3.8	1470	2	US-09-062-890-3	Sequence 3, Appli
c	192	8	3.8	1470	2	US-09-062-890-5	Sequence 5, Appli
c	193	8	3.8	1470	2	US-09-062-890-7	Sequence 7, Appli
c	194	8	3.8	1470	2	US-09-062-890-9	Sequence 9, Appli
c	195	8	3.8	1470	2	US-09-062-890-11	Sequence 11, Appl
c	196	8	3.8	1470	2	US-09-062-890-13	Sequence 13, Appl
c	197	8	3.8	1470	2	US-09-062-890-15	Sequence 15, Appl
c	198	8	3.8	1470	2	US-09-062-890-17	Sequence 17, Appl
c	199	8	3.8	1470	2	US-09-062-890-19	Sequence 19, Appl
c	200	8	3.8	1470	2	US-09-062-890-21	Sequence 21, Appl
c	201	8	3.8	1470	2	US-09-062-890-23	Sequence 23, Appl
c	202	8	3.8	1470	2	US-09-062-890-25	Sequence 25, Appl
c	203	8	3.8	1470	2	US-09-062-890-27	Sequence 27, Appl
c	204	8	3.8	1470	2	US-09-062-890-29	Sequence 29, Appl
c	205	8	3.8	1470	2	US-09-062-890-31	Sequence 31, Appl
c	206	8	3.8	1470	2	US-09-062-890-33	Sequence 33, Appl
c	207	8	3.8	1470	2	US-09-062-890-35	Sequence 35, Appl
c	208	8	3.8	1470	2	US-09-062-890-37	Sequence 37, Appl
c	209	8	3.8	1635	3	US-09-252-991A-3303	Sequence 3303, Ap
	210	8	3.8	1662	3	US-09-771-161A-61	Sequence 61, Appl
	211	8	3.8	1704	3	US-09-489-039A-3097	Sequence 3097, Ap
c	212	8	3.8	1713	3	US-09-489-039A-2250	Sequence 2250, Ap
	213	8	3.8	1742	3	US-08-466-368-3	Sequence 3, Appli
	214	8	3.8	1742	3	US-09-517-605-7	Sequence 7, Appli
	215	8	3.8	1742	3	US-08-470-998-1	Sequence 1, Appli
	216	8	3.8	1742	3	US-09-023-655-1013	Sequence 1013, Ap
	217	8	3.8	1742	3	US-08-328-500-8	Sequence 8, Appl
c	218	8	3.8	1757	3	US-09-799-451-933	Sequence 933, App
	219	8	3.8	1946	3	US-09-949-016-4138	Sequence 4138, Ap
	220	8	3.8	1977	3	US-09-614-891-5	Sequence 5, Appli
c	221	8	3.8	2016	3	US-09-252-991A-7070	Sequence 7070, Ap
	222	8	3.8	2091	2	US-08-286-856C-1	Sequence 1, Appli
	223	8	3.8	2091	2	US-08-472-831-1	Sequence 1, Appli
	224	8	3.8	2153	2	US-08-577-492-31	Sequence 31, Appl
	225	8	3.8	2153	3	US-09-079-630-31	Sequence 31, Appl
c	226	8	3.8	2195	3	US-09-771-161A-60	Sequence 60, Appl
c	227	8	3.8	2226	3	US-09-252-991A-14877	Sequence 14877, A

Left column

Flag	No.			Length	Accession	Sequence
c	228	3.8	8	2259	US-09-252-991A-13396	Sequence 13396, A
	229	3.8	8	2268	US-09-252-991A-14487	Sequence 14487, A
	230	3.8	8	2388	US-10-029-180-51	Sequence 51, Appl
	231	3.8	8	2422	US-09-949-016-3654	Sequence 3654, Ap
	232	3.8	8	2455	US-08-073-807A-1	Sequence 1, Appli
	233	3.8	8	2511	US-08-422-699A-8	Sequence 8, Appli
	234	3.8	8	2511	US-08-422-706B-8	Sequence 516, Appli
	235	3.8	8	2552	US-10-104-047-516	Sequence 516, Appli
c	236	3.8	8	2691	US-09-489-039A-3091	Sequence 3091, Ap
	237	3.8	8	2726	US-08-422-699A-12	Sequence 12, Appl
	238	3.8	8	2726	US-08-422-706B-12	Sequence 12, Appl
	239	3.8	8	2748	US-09-489-039A-1078	Sequence 1078, Ap
c	240	3.8	8	2914	US-10-104-047-1009	Sequence 1009, Ap
c	241	3.8	8	3222	US-09-252-991A-13746	Sequence 13746, A
	242	3.8	8	3323	US-08-422-699A-10	Sequence 10, Appl
	243	3.8	8	3323	US-08-422-706B-10	Sequence 10, Appl
	244	3.8	8	3387	US-09-194-640A-2	Sequence 2, Appli
c	245	3.8	8	3387	US-10-162-160-2	Sequence 2, Appli
c	246	3.8	8	3418	US-09-949-016-2090	Sequence 2090, Ap
	247	3.8	8	3498	US-10-104-047-952	Sequence 952, App
	248	3.8	8	3538	US-09-949-016-2767	Sequence 2767, Ap
c	249	3.8	8	3584	US-10-104-047-538	Sequence 538, App
	250	3.8	8	3868	US-09-357-070-1	Sequence 1, Appli
	251	3.8	8	3868	US-09-949-016-709	Sequence 709, App
	252	3.8	8	4270	US-09-949-016-12549	Sequence 12549, A
	253	3.8	8	4336	US-09-949-016-13959	Sequence 13959, A
	254	3.8	8	4406	US-10-104-047-979	Sequence 979, App
c	255	3.8	8	4797	US-09-643-597-134	Sequence 134, App
c	256	3.8	8	4797	US-09-480-884A-134	Sequence 134, App
c	257	3.8	8	4797	US-09-542-615A-134	Sequence 134, App
c	258	3.8	8	4797	US-09-606-421B-134	Sequence 134, App
c	259	3.8	8	4797	US-09-221-107-134	Sequence 29, Appl
c	260	3.8	8	4797	US-09-466-396A-134	Sequence 134, App
c	261	3.8	8	4797	US-09-476-496A-134	Sequence 134, App
c	262	3.8	8	4797	US-09-630-940B-134	Sequence 134, App
c	263	3.8	8	4797	US-09-285-479-134	Sequence 134, App
c	264	3.8	8	4797	US-10-007-700-134	Sequence 134, App
c	265	3.8	8	5056	US-09-949-016-1463	Sequence 1463, Ap
c	266	3.8	8	5065	US-09-949-016-358	Sequence 358, App
c	267	3.8	8	5176	US-08-654-482-13	Sequence 13, Appl
c	268	3.8	8	5176	US-09-585-023-13	Sequence 13, Appl
	269	3.8	8	5220	US-08-777-405A-1	Sequence 1, Appli
	270	3.8	8	5220	US-08-977-871A-1	Sequence 1, Appli
	271	3.8	8	5220	US-09-225-951-1	Sequence 1, Appli
	272	3.8	8	5220	US-09-841-341-1	Sequence 1, Appli
	273	3.8	8	5220	US-10-027-591-1	Sequence 1, Appli
	274	3.8	8	5220	US-10-337-192-1	Sequence 1, Appli
	275	3.8	8	5220	US-10-697-912-1	Sequence 1, Appli
	276	3.8	8	5325	US-09-919-172-75	Sequence 75, Appl
	277	3.8	8	6145	US-09-949-016-4619	Sequence 4619, Ap
	278	3.8	8	6655	US-09-902-540-777	Sequence 777, App
	279	3.8	8	6855	US-09-252-991A-6976	Sequence 6976, Ap
	280	3.8	8	7874	US-09-780-175-96	Sequence 96, Appl
	281	3.8	8	8341	US-08-737-825-1	Sequence 1, Appli
	282	3.8	8	9046	US-08-227-536-1	Sequence 1, Appli
	283	3.8	8	9046	PCT-US95-04682-1	Sequence 3, Appli
	284	3.8	8	9308	US-09-949-016-13107	Sequence 13107, A
	285	3.8	8	9681	US-09-949-016-12753	Sequence 12753, A
	286	3.8	8	9681	US-09-949-016-14849	Sequence 14849, A
	287	3.8	8	10655	US-09-949-016-13868	Sequence 13868, A
	288	3.8	8	11613	US-08-484-044-10	Sequence 10, Appl
c	289	3.8	8	14483	US-10-164-230-3	Sequence 3, Appli
c	290	3.8	8	14875	US-09-949-016-13929	Sequence 13929, A
c	291	3.8	8	16367	US-09-949-016-14509	Sequence 14509, A
	292	3.8	8	19085	US-09-949-016-16361	Sequence 16361, A
	293	3.8	8	19761	US-09-949-016-14236	Sequence 14236, A
	294	3.8	8	19762	US-09-949-016-12814	Sequence 12814, A
	295	3.8	8	22611	US-09-949-016-12379	Sequence 12379, A
	296	3.8	8	22611	US-09-949-016-15883	Sequence 15883, A
	297	3.8	8	23292	US-09-949-016-12100	Sequence 12100, A
c	298	3.8	8	23292	US-09-949-016-13205	Sequence 13205, A
c	299	3.8	8	23927	US-09-949-016-14284	Sequence 14284, A
c	300	3.8	8	24020	US-09-949-016-17353	Sequence 17353, A

Right column

Flag	No.	Ct			Length	Accession	Sequence
	301		8	3.8	25370	US-09-949-016-12109	Sequence 12109, A
	302		8	3.8	25375	US-09-949-016-15880	Sequence 15880, A
	303		8	3.8	26174	US-09-949-016-15229	Sequence 15229, A
c	304		8	3.8	27578	US-09-949-016-12167	Sequence 12167, A
c	305		8	3.8	27579	US-09-949-016-15005	Sequence 15005, A
	306		8	3.8	28438	US-09-820-790B-3	Sequence 3, Appl
	307		8	3.8	28791	US-09-949-016-15396	Sequence 15396, A
	308		8	3.8	29172	US-09-949-016-15520	Sequence 15520, A
	309		8	3.8	29433	US-09-949-016-15740	Sequence 15740, A
c	310		8	3.8	31797	US-09-949-016-17188	Sequence 17188, A
c	311		8	3.8	32099	US-09-949-016-16562	Sequence 16562, A
c	312		8	3.8	35630	US-09-949-016-13832	Sequence 13832, A
c	313		8	3.8	38705	US-09-949-016-13883	Sequence 13883, A
c	314		8	3.8	38705	US-09-949-016-13884	Sequence 13884, A
	315		8	3.8	41523	US-09-949-016-11932	Sequence 11932, A
	316		8	3.8	41522	US-09-949-016-15764	Sequence 15764, A
c	317		8	3.8	54484	US-09-902-540-1272	Sequence 1272, Ap
c	318		8	3.8	68452	US-09-949-016-13305	Sequence 13305, A
	319		8	3.8	87350	US-08-781-891-79	Sequence 79, Appl
c	320		8	3.8	87350	US-09-618-166-79	Sequence 79, Appl
	321		8	3.8	87543	US-09-791-211-3	Sequence 3, Appli
	322		8	3.8	87780	US-09-949-016-17011	Sequence 17011, A
	323		8	3.8	88950	US-09-949-016-17150	Sequence 17150, A
	324		8	3.8	90428	US-09-949-016-12564	Sequence 12564, A
	325		8	3.8	94095	US-09-949-016-14389	Sequence 14389, A
	326		8	3.8	109159	US-09-949-016-14170	Sequence 14170, A
	327		8	3.8	109159	US-09-949-016-14169	Sequence 14169, A
	328		8	3.8	112874	US-09-949-016-13180	Sequence 13180, A
	329		8	3.8	120727	US-09-949-016-15788	Sequence 15788, A
	330		8	3.8	120727	US-09-949-016-15787	Sequence 15787, A
	331		8	3.8	133719	US-09-949-016-15092	Sequence 15092, A
	332		8	3.8	143776	US-09-949-001-29	Sequence 29, Appl
	333		8	3.8	144034	US-09-949-016-35	Sequence 35, Appl
	334		8	3.8	178883	US-09-949-016-12733	Sequence 12733, A
	335		8	3.8	178884	US-09-949-016-13039	Sequence 13039, A
	336		8	3.8	191433	US-09-949-016-16144	Sequence 16144, A
	337		8	3.8	245286	US-09-949-016-15497	Sequence 15497, A
	338		8	3.8	263693	US-09-949-016-12386	Sequence 12386, A
	339		8	3.8	263694	US-09-949-016-16915	Sequence 16915, A
	340		8	3.8	325791	US-09-768-185A-1	Sequence 1, Appl
	341		8	3.8	340380	US-09-949-016-14179	Sequence 14179, A
	342	3		3.8	1664976	US-08-916-421B-1	Sequence 1, Appli
	343	3		3.8	1664976	US-09-692-570-1	Sequence 1, Appli
	344	3		3.8	4403765	US-09-103-840A-2	Sequence 2, Appli
	345	3		3.8	4403765	US-09-103-840A-2	Sequence 2, Appli
	346	3		3.8	4411529	US-09-103-840A-1	Sequence 1, Appli
	347	3		3.8	4411529	US-09-103-840A-1	Sequence 1, Appli
	348	3	7	3.3	37	US-09-024-020B-34	Sequence 34, App
	349	3	7	3.3	37	US-09-425-043-34	Sequence 34, App
	350	2	7	3.3	38	US-09-375-975-35	Sequence 35, App
	351	3	7	3.3	38	US-10-461-069-35	Sequence 35, App
	352	2	7	3.3	48	US-08-221-817-24	Sequence 24, App
	353	2	7	3.3	48	US-08-454-439-24	Sequence 24, App
	354	6	7	3.3	48	PCT-US94-10487-24	Sequence 24, App
	355	2	7	3.3	60	US-08-484-192-152	Sequence 152, App
	356	3	7	3.3	68	US-09-382-552-96	Sequence 96, Appl
	357	3	7	3.3	86	US-08-464-700-30	Sequence 30, Appl
	358	3	7	3.3	90	US-08-556-978B-93	Sequence 93, Appl
	359	3	7	3.3	91	US-08-556-978B-94	Sequence 94, Appl
c	360	2	7	3.3	96	US-09-270-767-27877	Sequence 27877,
	361	3	7	3.3	142	US-08-464-700-5	Sequence 5, Appl
	362	3	7	3.3	150	US-09-313-294A-589	Sequence 589, App
c	363	3	7	3.3	150	US-08-886-280-5	Sequence 5, Appl
	364	3	7	3.3	156	US-08-870-370-6	Sequence 6, Appl
	365	3	7	3.3	158	US-09-513-999C-8511	Sequence 8511, Ap
	366	3	7	3.3	160	US-09-513-999C-14643	Sequence 14643, A
	367	3	7	3.3	184	US-09-513-999C-20652	Sequence 20652, A
c	368	3	7	3.3	184	US-08-637-759B-16	Sequence 16, Appl
c	369	3	7	3.3	184	US-08-871-355A-16	Sequence 16, Appl
c	370	3	7	3.3	185	US-09-201-945-16	Sequence 16, Appl
	371	3	7	3.3	185	US-09-513-999C-8436	Sequence 8436, Ap
c	372	3	7	3.3	195	US-09-328-352-2392	Sequence 2392, Ap
	373	3	7	3.3	196	US-09-513-999C-11916	Sequence 11916, A

	#	score		Accession	n	Sequence
c	374	3.3	7	US-09-270-767-29003	3	Sequence 29003, A
c	375	3.3	7	US-09-543-681A-1154	3	Sequence 1154, Ap
c	376	3.3	7	US-10-237-551-70	3	Sequence 70, Appl
c	377	3.3	7	US-09-513-999C-21077	3	Sequence 21077, A
	378	3.3	7	US-09-016-434-519	3	Sequence 519, App
c	379	3.3	7	US-09-270-767-785	3	Sequence 785, App
c	380	3.3	7	US-09-270-767-16067	3	Sequence 16067, A
c	381	3.3	7	US-09-016-434-598	3	Sequence 598, App
	382	3.3	7	US-09-145-650D-15	2	Sequence 15, Appl
	383	3.3	7	US-09-513-999C-3448	3	Sequence 3448, Ap
	384	3.3	7	US-09-328-352-2813	3	Sequence 2813, Ap
	385	3.3	7	US-09-382-552-83	3	Sequence 83, Appl
c	386	3.3	7	US-09-902-540-6261	3	Sequence 6261, Ap
c	387	3.3	7	US-09-513-999C-28328	3	Sequence 28328, A
	388	3.3	7	US-09-489-039A-2539	3	Sequence 2539, Ap
	389	3.3	7	US-09-313-294A-3448	3	Sequence 3448, Ap
c	390	3.3	7	US-09-016-434-332	3	Sequence 332, App
c	391	3.3	7	US-09-313-294A-3404	3	Sequence 3404, Ap
	392	3.3	7	US-09-328-352-2447	3	Sequence 2447, Ap
	393	3.3	7	US-09-358-321C-3	3	Sequence 3, Appli
c	394	3.3	7	US-09-313-294A-3399	3	Sequence 3399, Ap
	395	3.3	7	US-09-313-294A-5329	3	Sequence 5329, Ap
	396	3.3	7	US-09-270-767-29663	3	Sequence 29663, A
	397	3.3	7	US-09-270-767-29457	3	Sequence 29457, A
c	398		9	Patent No. 5191064-4		Patent No. 5191064
c	399	3.3	7	US-09-313-294A-4345	3	Sequence 4345, Ap
c	400	3.3	7	US-09-313-294A-4515	3	Sequence 4515, Ap
c	401	3.3	7	US-09-313-294A-6384	3	Sequence 6384, Ap
c	402	3.3	7	US-09-513-999C-8924	3	Sequence 8924, Ap
	403	3.3	7	US-09-513-999C-8692	3	Sequence 8692, Ap
	404	3.3	7	US-09-513-999C-13107	3	Sequence 13107, A
	405	3.3	7	US-09-513-999C-1232	3	Sequence 1232, Ap
	406	3.3	7	US-08-081-072-7	2	Sequence 7, Appli
	407	3.3	7	US-08-449-093A-7	3	Sequence 7, Appli
	408	3.3	7	US-09-902-540-2082	3	Sequence 2082, Ap
c	409	3.3	7	US-09-702-705-976	3	Sequence 976, App
	410	3.3	7	US-09-736-457-976	3	Sequence 976, App
	411	3.3	7	US-09-614-124B-976	3	Sequence 976, App
	412	3.3	7	US-09-671-325-976	3	Sequence 976, App
	413	3.3	7	US-09-658-824-976	3	Sequence 976, App
	414	3.3	7	US-10-017-754-976	3	Sequence 976, App
	415	3.3	7	US-09-651-563-976	3	Sequence 976, App
c	416	3.3	7	US-09-489-039A-4215	3	Sequence 4215, Ap
	417	3.3	7	US-09-902-540-6115	3	Sequence 6115, Ap
	418	3.3	7	US-09-489-039A-2927	3	Sequence 2927, Ap
	419	3.3	7	US-09-008-892-9	3	Sequence 9, Appli
c	420	3.3	7	US-09-513-999C-2052	3	Sequence 2052, Ap
	421	3.3	7	US-09-513-999C-21483	3	Sequence 21483, A
	422	3.3	7	US-08-072-574-3	3	Sequence 3, Appli
	423	3.3	7	US-08-486-270-3	3	Sequence 3, Appli
	424	3.3	7	US-08-367-264-3	3	Sequence 3, Appli
	425	3.3	7	US-09-153-757-3	3	Sequence 3, Appli
	426	3.3	7	US-09-459-715-3	3	Sequence 3, Appli
c	427	3.3	7	US-09-248-796A-13456	3	Sequence 13456, A
c	428	3.3	7	US-09-513-999C-25678	3	Sequence 25678, A
c	429	3.3	7	US-09-270-767-28516	3	Sequence 28516, A
c	430	3.3	7	US-09-949-002-1612	3	Sequence 1612, Ap
c	431	3.3	7	US-09-902-540-6977	3	Sequence 6977, Ap
c	432	3.3	7	US-09-902-540-7487	3	Sequence 7487, Ap
c	433	3.3	7	US-09-640-011A-499	3	Sequence 499, App
c	434	3.3	7	US-09-489-039A-5014	3	Sequence 5014, Ap
	435	3.3	7	US-09-621-976-15471	3	Sequence 15471, A
	436	3.3	7	US-09-643-597-215	3	Sequence 215, App
	437	3.3	7	US-09-480-884A-215	3	Sequence 215, App
	438	3.3	7	US-09-542-615A-215	3	Sequence 215, App
	439	3.3	7	US-09-606-421B-215	3	Sequence 215, App
	440	3.3	7	US-09-466-396A-215	3	Sequence 215, App
	441	3.3	7	US-09-476-496A-215	3	Sequence 215, App
	442	3.3	7	US-09-630-940B-215	3	Sequence 215, App
	443	3.3	7	US-10-007-700-215	3	Sequence 215, App
	444	3.3	7	US-09-621-976-13806	3	Sequence 13806, A
	445	3.3	7	US-09-345-882-5	3	Sequence 5, Appli
	446	3.3	7	US-10-071-179-5	3	Sequence 5, Appli

	#	score		num	n	Accession	Sequence
c	447	3.3	7	397	2	US-08-592-126-106	Sequence 106, App
c	448	3.3	7	397	3	US-09-168-595-106	Sequence 106, App
c	449	3.3	7	398	3	US-09-513-999C-347	Sequence 347, App
c	450	3.3	7	404	3	US-09-513-999C-32191	Sequence 32191, A
c	451	3.3	7	409	3	US-09-513-999C-28527	Sequence 28527, A
c	452	3.3	7	410	3	US-09-854-133-400	Sequence 400, App
c	453	3.3	7	416	3	US-09-513-999C-23414	Sequence 23414, A
	454	3.3	7	417	3	US-09-252-991A-873	Sequence 873, App
	455	3.3	7	421	3	US-09-252-991A-2706	Sequence 2706, Ap
	456	3.3	7	421	3	US-09-270-767-8659	Sequence 8659, Ap
	457	3.3	7	422	3	US-09-270-767-23941	Sequence 23941, A
	458	3.3	7	423	3	US-09-640-211A-1524	Sequence 1524, Ap
	459	3.3	7	432	3	US-09-513-999C-13328	Sequence 13328, A
	460	3.3	7	432	3	US-09-252-991A-1831	Sequence 1831, Ap
	461	3.3	7	433	3	US-09-252-991A-6126	Sequence 6126, Ap
	462	3.3	7	434	3	US-09-949-016-69334	Sequence 69334, A
	463	3.3	7	434	3	US-09-621-976-14295	Sequence 14295, A
	464	3.3	7	435	3	US-09-621-976-13280	Sequence 13280, A
	465	3.3	7	435	3	US-10-021-338A-71	Sequence 71, Appli
c	466	3.3	7	438	3	US-08-692-922-3	Sequence 3, Appli
	467	3.3	7	441	3	US-09-252-991A-3947	Sequence 3947, Ap
	468	3.3	7	443	3	US-09-328-352-1472	Sequence 1472, Ap
c	469	3.3	7	447	3	US-08-914-375C-51	Sequence 51, Appl
c	470	3.3	7	449	3	US-09-621-976-14389	Sequence 14389, A
	471	3.3	7	449	3	US-09-252-991A-9576	Sequence 9576, Ap
	472	3.3	7	449	3	US-08-478-507-13	Sequence 13, Appl
	473	3.3	7	450	3	US-09-128-275A-13	Sequence 13, Appl
	474	3.3	7	452	3	US-09-553-427-13	Sequence 13, Appl
	475	3.3	7	456	3	US-09-513-999C-12530	Sequence 12530, A
	476	3.3	7	462	3	US-09-976-594-1094	Sequence 1094, Ap
c	477	3.3	7	464	3	US-09-513-999C-325	Sequence 325, App
	478	3.3	7	465	3	US-09-252-991A-6955	Sequence 6955, Ap
c	479	3.3	7	468	3	US-09-270-767-11917	Sequence 11917, A
c	480	3.3	7	480	3	US-09-513-999C-465	Sequence 465, App
c	481	3.3	7	481	3	US-09-252-991A-13661	Sequence 13661, A
c	482	3.3	7	482	3	US-09-252-991A-894	Sequence 894, App
c	483	3.3	7	484	3	US-09-489-039A-2985	Sequence 2985, Ap
c	484	3.3	7	485	3	US-09-471-276-437	Sequence 437, App
c	485	3.3	7	486	3	US-09-513-999C-3953	Sequence 3953, Ap
c	486	3.3	7	487	2	US-08-812-828-2	Sequence 2, Appli
c	487	3.3	7	488	2	US-09-244-192-2	Sequence 2, Appli
c	488	3.3	7	488	3	US-09-489-039A-222	Sequence 222, App
c	489	3.3	7	489	3	US-09-949-016-2887	Sequence 2887, App
	490	3.3	7	490	3	US-09-621-976-1099	Sequence 1099, Ap
	491	3.3	7	491	3	US-09-252-991A-15526	Sequence 15526, A
	492	3.3	7	492	3	US-09-270-767-7386	Sequence 7386, App
	493	3.3	7	493	3	US-09-270-767-22668	Sequence 22668, A
	494	3.3	7	494	3	US-09-270-767-28588	Sequence 28588, A
	495	3.3	7	495	3	US-09-270-767-13118	Sequence 13118, A
c	496	3.3	7	496	3	US-09-252-991A-3081	Sequence 3081, Ap
c	497	3.3	7	501	3	US-09-621-976-1974	Sequence 1974, Ap
c	498	3.3	7	502	3	US-09-270-767-5718	Sequence 5718, Ap
c	499	3.3	7	502	3	US-09-270-767-21000	Sequence 21000, A
c	500	3.3	7	502	3	US-09-258-031C-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-10-104-047-799
; Sequence 799, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: H1-A0105
; FILE REFERENCE: 6943241e1 full length cDNA
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 799

```
;    LENGTH: 1785
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-104-047-799

Alignment Scores:
Pred. No.:                    4.26e-95          Length:            1785
Score:                        115.00            Matches:           211
Percent Similarity:           99.1%             Conservative:      0
Best Local Similarity:        99.1%             Mismatches:        1
Query Match:                  54.2%             Indels:            2
DB:                                             Gaps:              0

US-09-989-890-238 (1-212) x US-10-104-047-799 (1-1785)

Qy      1 SerProHisGlnAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 AGCCCCCACCAAGCCGCCCGACCCGTAGACCAGACCCCAAGGACCCTGGCCACCATGGGC 252

Qy     21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuProSerProProAla 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 CAGAGAGCAGATTACCTTCATCTCTGGCTCTCGTCGAGCCGGCCCTTGAGTCCCCACCTGCT 312

Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313 GCCTGCTCTGGCGACCCTGGGTGTGGAGGTGGTGCCGGAGGTGCCTGCCTTCCGCCGGCT 372

Qy     61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373 GCCGGGAATTGCCTCCAGCGCTGTGGAGCCTGTGTGCGGGGATGCAGCCCCTGCCTGTCTA 432

Qy     81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433 CTGAGGACTCCACTGAGGGACTGCTGAAGCCAACTGGCCAACGAGGACACAATGGAGTGC 492

Qy    101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGly-SerSerGlnPr 120
            ||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    493 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGATGGCCAGCAG-CTCAAGTCAACC 551

Qy    120 oTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProTh 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    552 ATGGGCAGCAGCTTCAGCTACCCCATGTTAAGCTCAAAGGCATCCCTGTATCCCTAC 611

Qy    140 rArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProl 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    612 CCGAGGGGCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGAT 671

Qy    160 eProHisProCysAspThrAlaCysProLeuProValLeuValAlaProAr 180
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db    672 CCCCCACCCATGCGACACAGCCCTGCCCCAGCACCTTTGCCAGTCCTCGTGGCGTCCGAG 731

Qy    180 gSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCy 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    732 GAGTACTATTCTTTCCATGATCCGACCTGACCCTGCCGCCGGAGATGGGCAGTGGCTCCATG 791

Qy    200 sArgAlaGluLysMetCysSerSerSerArgSer 212
            ||||||||||||||||||||||||||||||
Db    792 TCGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 828

RESULT 2
US-09-949-016-17032
; Sequence 17032, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;    APPLICANT: VENTER, J. Craig et al.
;    TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;    TITLE OF INVENTION:   WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;    FILE REFERENCE: CL001307
;    CURRENT APPLICATION NUMBER: US/09/949,016
;    CURRENT FILING DATE: 2000-04-14
;    PRIOR APPLICATION NUMBER: 60/241,755
;    PRIOR FILING DATE: 2000-10-20
;    PRIOR APPLICATION NUMBER: 60/237,768
```

```
;    PRIOR FILING DATE: 2000-10-03
;    PRIOR APPLICATION NUMBER: 60/231,498
;    PRIOR FILING DATE: 2000-09-08
;    NUMBER OF SEQ ID NOS: 207012
;    SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17032
;    LENGTH: 194937
;    TYPE: DNA
;    ORGANISM: Human
US-09-949-016-17032

Alignment Scores:
Pred. No.:                    114               Length:            194937
Score:                        11.00             Matches:           11
Percent Similarity:           100.0%            Conservative:      0
Best Local Similarity:        100.0%            Mismatches:        0
Query Match:                  5.2%              Indels:            0
DB:                                             Gaps:              0

US-09-989-890-238 (1-212) x US-09-949-016-17032 (1-194937)

Qy     24 LeuProSerSerLeuAlaLeuLeuSerArgPro 34
            |||||||||||||||||||||||||||||||||
Db 182186 CTTCCATCCAGCCTCGCTCTCCTGAGCCGCCCC 182218

RESULT 3
US-09-949-016-17033
; Sequence 17033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;    APPLICANT: VENTER, J. Craig et al.
;    TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;    TITLE OF INVENTION:   WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;    FILE REFERENCE: CL001307
;    CURRENT APPLICATION NUMBER: US/09/949,016
;    CURRENT FILING DATE: 2000-04-14
;    PRIOR APPLICATION NUMBER: 60/241,755
;    PRIOR FILING DATE: 2000-10-20
;    PRIOR APPLICATION NUMBER: 60/237,768
;    PRIOR FILING DATE: 2000-10-03
;    PRIOR APPLICATION NUMBER: 60/231,498
;    PRIOR FILING DATE: 2000-09-08
;    NUMBER OF SEQ ID NOS: 207012
;    SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17033
;    LENGTH: 194937
;    TYPE: DNA
;    ORGANISM: Human
US-09-949-016-17033

Alignment Scores:
Pred. No.:                    114               Length:            194937
Score:                        11.00             Matches:           11
Percent Similarity:           100.0%            Conservative:      0
Best Local Similarity:        100.0%            Mismatches:        0
Query Match:                  5.2%              Indels:            0
DB:                                             Gaps:              0

US-09-989-890-238 (1-212) x US-09-949-016-17033 (1-194937)

Qy     24 LeuProSerSerLeuAlaLeuLeuSerArgPro 34
            |||||||||||||||||||||||||||||||||
Db 182186 CTTCCATCCAGCCTCGCTCTCCTGAGCCGCCCC 182218

RESULT 4
US-09-949-016-12803
; Sequence 12803, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;    APPLICANT: VENTER, J. Craig et al.
;    TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;    TITLE OF INVENTION:   WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;    FILE REFERENCE: CL001307
```

```
;      CURRENT APPLICATION NUMBER: US/09/949,016
;,     CURRENT FILING DATE: 2000-04-14
;,     PRIOR APPLICATION NUMBER: 60/241,755
;,     PRIOR FILING DATE: 2000-10-20
;,     PRIOR APPLICATION NUMBER: 60/237,768
;,     PRIOR FILING DATE: 2000-10-03
;,     PRIOR APPLICATION NUMBER: 60/231,498
;,     PRIOR FILING DATE: 2000-09-08
;      NUMBER OF SEQ ID NOS: 207012
;      SOFTWARE: FastSEQ for Windows Version 4.0
;      SEQ ID NO 12803
;      LENGTH: 31861
;      TYPE: DNA
;      ORGANISM: Human
US-09-949-016-12803

Alignment Scores:
Pred. No.:                      177                     Length:              31861
Score:                          10.00                   Matches:             10
Percent Similarity:             100.0%                  Conservative:        0
Best Local Similarity:          100.0%                  Mismatches:          0
Query Match:                    4.7%                    Indels:              0
DB:                             3                       Gaps:                0

US-09-989-890-238 (1-212) x US-09-949-016-12803 (1-31861)

Qy        23 AlaLeuProSerSerAlaAlaLeuLeuSer 32
             ||||||||||||||||||||||||||||||
Db      1579 GCCCTGCCCTCCAGCCTGGCTCTCCTTTCC 1608

RESULT 5
US-09-949-016-13967
; Sequence 13967, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;,     APPLICANT: VENTER, J. Craig et al.
;,     TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;,     TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;,     FILE REFERENCE: CL001307
;,     CURRENT APPLICATION NUMBER: US/09/949,016
;,     CURRENT FILING DATE: 2000-04-14
;,     PRIOR APPLICATION NUMBER: 60/241,755
;,     PRIOR FILING DATE: 2000-10-20
;,     PRIOR APPLICATION NUMBER: 60/237,768
;,     PRIOR FILING DATE: 2000-10-03
;,     PRIOR APPLICATION NUMBER: 60/231,498
;,     PRIOR FILING DATE: 2000-09-08
;      NUMBER OF SEQ ID NOS: 207012
;      SOFTWARE: FastSEQ for Windows Version 4.0
;      SEQ ID NO 13967
;      LENGTH: 31861
;      TYPE: DNA
;      ORGANISM: Human
US-09-949-016-13967

Alignment Scores:
Pred. No.:                      177                     Length:              31861
Score:                          10.00                   Matches:             10
Percent Similarity:             100.0%                  Conservative:        0
Best Local Similarity:          100.0%                  Mismatches:          0
Query Match:                    4.7%                    Indels:              0
DB:                             3                       Gaps:                0

US-09-989-890-238 (1-212) x US-09-949-016-13967 (1-31861)

Qy        23 AlaLeuProSerSerAlaAlaLeuLeuSer 32
             ||||||||||||||||||||||||||||||
Db      1579 GCCCTGCCCTCCAGCCTGGCTCTCCTTTCC 1608

RESULT 6
US-09-949-016-30920/c
; Sequence 30920, Application US/09949016
; Patent No. 6812339
```

```
;      GENERAL INFORMATION:
;,     APPLICANT: VENTER, J. Craig et al.
;,     TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;,     TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;,     FILE REFERENCE: CL001307
;,     CURRENT APPLICATION NUMBER: US/09/949,016
;,     CURRENT FILING DATE: 2000-04-14
;,     PRIOR APPLICATION NUMBER: 60/241,755
;,     PRIOR FILING DATE: 2000-10-20
;,     PRIOR APPLICATION NUMBER: 60/237,768
;,     PRIOR FILING DATE: 2000-10-03
;,     PRIOR APPLICATION NUMBER: 60/231,498
;,     PRIOR FILING DATE: 2000-09-08
;      NUMBER OF SEQ ID NOS: 207012
;      SOFTWARE: FastSEQ for Windows Version 4.0
;      SEQ ID NO 30920
;      LENGTH: 601
;      TYPE: DNA
;      ORGANISM: Human
US-09-949-016-30920

Alignment Scores:
Pred. No.:                      38.4                    Length:              601
Score:                          9.00                    Matches:             9
Percent Similarity:             100.0%                  Conservative:        0
Best Local Similarity:          100.0%                  Mismatches:          0
Query Match:                    4.2%                    Indels:              0
DB:                             3                       Gaps:                0

US-09-989-890-238 (1-212) x US-09-949-016-30920 (1-601)

Qy        25 ProSerSerLeuAlaLeuLeuSerArg 33
             |||||||||||||||||||||||||||
Db       291 CCGAGCTCCCTGGCTCTACTATCCAGG 265

RESULT 7
US-09-949-016-30921/c
; Sequence 30921, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;,     APPLICANT: VENTER, J. Craig et al.
;,     TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;,     TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;,     FILE REFERENCE: CL001307
;,     CURRENT APPLICATION NUMBER: US/09/949,016
;,     CURRENT FILING DATE: 2000-04-14
;,     PRIOR APPLICATION NUMBER: 60/241,755
;,     PRIOR FILING DATE: 2000-10-20
;,     PRIOR APPLICATION NUMBER: 60/237,768
;,     PRIOR FILING DATE: 2000-10-03
;,     PRIOR APPLICATION NUMBER: 60/231,498
;,     PRIOR FILING DATE: 2000-09-08
;      NUMBER OF SEQ ID NOS: 207012
;      SOFTWARE: FastSEQ for Windows Version 4.0
;      SEQ ID NO 30921
;      LENGTH: 601
;      TYPE: DNA
;      ORGANISM: Human
US-09-949-016-30921

Alignment Scores:
Pred. No.:                      38.4                    Length:              601
Score:                          9.00                    Matches:             9
Percent Similarity:             100.0%                  Conservative:        0
Best Local Similarity:          100.0%                  Mismatches:          0
Query Match:                    4.2%                    Indels:              0
DB:                             3                       Gaps:                0

US-09-989-890-238 (1-212) x US-09-949-016-30921 (1-601)

Qy        25 ProSerSerLeuAlaLeuLeuSerArg 33
             |||||||||||||||||||||||||||
Db       183 CCGAGCTCCCTGGCTCTACTATCCAGG 157
```

RESULT 8
US-09-949-016-30922/c
; Sequence 30922, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30922
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-30922

Alignment Scores:
Pred. No.:   Length: 601
Score: 38.4   Matches: 9
                9.00   Conservative: 0
Percent Similarity: 100.0%   Mismatches: 0
Best Local Similarity: 100.0%   Indels: 0
Query Match: 4.2%   Gaps: 0
DB: 3

US-09-989-890-238 (1-212) x US-09-949-016-30922 (1-601)

Qy   25 ProSerLeuAlaLeuLeuSerArg 33
           |||||||||||||||||||||||||
Db   182 CCGAGCTCCCTGGCTCTACTATCCAGG 156

RESULT 9
US-09-949-016-30923/c
; Sequence 30923, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30923
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-30923

Alignment Scores:
Pred. No.:   Length: 601
Score: 38.4   Matches: 9
                9.00   Conservative: 0
Percent Similarity: 100.0%   Mismatches: 0
Best Local Similarity: 100.0%   Indels: 0
Query Match: 4.2%   Gaps: 0
DB: 3

US-09-989-890-238 (1-212) x US-09-949-016-30923 (1-601)

Qy   25 ProSerLeuAlaLeuLeuSerArg 33
           |||||||||||||||||||||||||
Db   88 CCGAGCTCCCTGGCTCTACTATCCAGG 62

RESULT 10
US-09-949-016-30924/c
; Sequence 30924, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30924
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-30924

Alignment Scores:
Pred. No.:   Length: 601
Score: 38.4   Matches: 9
                9.00   Conservative: 0
Percent Similarity: 100.0%   Mismatches: 0
Best Local Similarity: 100.0%   Indels: 0
Query Match: 4.2%   Gaps: 0
DB: 3

US-09-989-890-238 (1-212) x US-09-949-016-30924 (1-601)

Qy   25 ProSerLeuAlaLeuLeuSerArg 33
           |||||||||||||||||||||||||
Db   70 CCGAGCTCCCTGGCTCTACTATCCAGG 44

RESULT 11
US-09-949-016-168951/c
; Sequence 168951, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168951
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-168951

Alignment Scores:
Pred. No.:   Length: 601
Score: 38.4   Matches: 9
                9.00

```
Percent Similarity:     100.0%       Conservative:    0
Best Local Similarity:  100.0%       Mismatches:      0
Query Match:            4.2%         Indels:          3
DB:                                  Gaps:            0

US-09-989-890-238 (1-212) x US-09-949-016-168951 (1-601)

Qy     25 ProSerSerLeuAlaLeuLeuSerArg 33
          |||||||||||||||||||||||
Db    291 CCGAGCTCCCTGGCTCTACTATCCAGG 265

RESULT 12
US-09-949-016-168952/c
; Sequence 168952, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE:  2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168952
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-168952

Alignment Scores:
Pred. No.:              38.4         Length:          601
Score:                  9.00         Matches:         9
Percent Similarity:     100.0%       Conservative:    0
Best Local Similarity:  100.0%       Mismatches:      0
Query Match:            4.2%         Indels:          0
DB:                                  Gaps:            0

US-09-989-890-238 (1-212) x US-09-949-016-168952 (1-601)

Qy     25 ProSerSerLeuAlaLeuLeuSerArg 33
          |||||||||||||||||||||||
Db    183 CCGAGCTCCCTGGCTCTACTATCCAGG 157

RESULT 13
US-09-949-016-168953/c
; Sequence 168953, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE:  2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168953
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
```

```
US-09-949-016-168953

Alignment Scores:
Pred. No.:              38.4         Length:          601
Score:                  9.00         Matches:         9
Percent Similarity:     100.0%       Conservative:    0
Best Local Similarity:  100.0%       Mismatches:      0
Query Match:            4.2%         Indels:          0
DB:                                  Gaps:            0

US-09-989-890-238 (1-212) x US-09-949-016-168953 (1-601)

Qy     25 ProSerLeuAlaLeuLeuSerArg 33
          ||||||||||||||||||||||
Db    182 CCGAGCTCCCTGGCTCTACTATCCAGG 156

RESULT 14
US-09-949-016-168954/c
; Sequence 168954, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE:  2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168954
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-168954

Alignment Scores:
Pred. No.:              38.4         Length:          601
Score:                  9.00         Matches:         9
Percent Similarity:     100.0%       Conservative:    0
Best Local Similarity:  100.0%       Mismatches:      0
Query Match:            4.2%         Indels:          0
DB:                                  Gaps:            0

US-09-989-890-238 (1-212) x US-09-949-016-168954 (1-601)

Qy     25 ProSerLeuAlaLeuLeuSerArg 33
          ||||||||||||||||||||||
Db     88 CCGAGCTCCCTGGCTCTACTATCCAGG 62

RESULT 15
US-09-949-016-168955/c
; Sequence 168955, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE:  2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 168955
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-168955

Alignment Scores:
Pred. No.:                        38.4      Length:           601
Score:                            9.00      Matches:          9
Percent Similarity:             100.0%      Conservative:     0
Best Local Similarity:          100.0%      Mismatches:       0
Query Match:                      4.2%      Indels:           0
DB:                                  3      Gaps:             0

US-09-989-890-238 (1-212) x US-09-949-016-168955 (1-601)

Qy         25 ProSerSerLeuAlaLeuLeuSerArg 33
              |||||||||||||||||||||||||||
Db         70 CCGAGCTCCCTGGCTCTACTATCCAGG 44


RESULT 16
US-09-902-540-8319/c
;  Sequence 8319, Application US/09902540
;  Patent No. 6833447
;  GENERAL INFORMATION:
;   APPLICANT: Goldman, Barry S.
;   APPLICANT: Hinkle, Gregory J.
;   APPLICANT: Slater, Steven C.
;   APPLICANT: Wiegand, Roger C.
;   TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;   FILE REFERENCE: 38-10(15849)B
;   CURRENT APPLICATION NUMBER: US/09/902,540
;   CURRENT FILING DATE: 2001-07-10
;   PRIOR APPLICATION NUMBER: 60/217,883
;   PRIOR FILING DATE: 2000-07-10
;   NUMBER OF SEQ ID NOS: 16825
;  SEQ ID NO 8319
;   LENGTH: 1377
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-8319

Alignment Scores:
Pred. No.:                        81.8      Length:          1377
Score:                            9.00      Matches:          9
Percent Similarity:             100.0%      Conservative:     0
Best Local Similarity:          100.0%      Mismatches:       0
Query Match:                      4.2%      Indels:           0
DB:                                  3      Gaps:             0

US-09-989-890-238 (1-212) x US-09-902-540-8319 (1-1377)

Qy         56 SerAlaSerAlaAlaGlyIleAla 64
              ||||||||||||||||||||||||
Db       1289 AGTGCCAGCGCCGGCTGCGGAATCGCT 1263


RESULT 17
US-09-902-540-9034/c
;  Sequence 9034, Application US/09902540
;  Patent No. 6833447
;  GENERAL INFORMATION:
;   APPLICANT: Goldman, Barry S.
;   APPLICANT: Hinkle, Gregory J.
;   APPLICANT: Slater, Steven C.
;   APPLICANT: Wiegand, Roger C.
;   TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;   FILE REFERENCE: 38-10(15849)B
;   CURRENT APPLICATION NUMBER: US/09/902,540
;   CURRENT FILING DATE: 2001-07-10
;   PRIOR APPLICATION NUMBER: 60/217,883
;   PRIOR FILING DATE: 2000-07-10
;   NUMBER OF SEQ ID NOS: 16825
;  SEQ ID NO 9034
;   LENGTH: 1716
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-9034

Alignment Scores:
Pred. No.:                         100      Length:          1716
Score:                            9.00      Matches:          9
Percent Similarity:             100.0%      Conservative:     0
Best Local Similarity:          100.0%      Mismatches:       0
Query Match:                      4.2%      Indels:           0
DB:                                  3      Gaps:             0

US-09-989-890-238 (1-212) x US-09-902-540-9034 (1-1716)

Qy         50 SerGlyAlaGlyLeuProSerAlaSer 58
              |||||||||||||||||||||||||||
Db       1342 TCAGGTGCTGGCCTTCCTTCAGCGAGT 1316


RESULT 18
US-09-902-540-868/c
;  Sequence 868, Application US/09902540
;  Patent No. 6833447
;  GENERAL INFORMATION:
;   APPLICANT: Goldman, Barry S.
;   APPLICANT: Hinkle, Gregory J.
;   APPLICANT: Slater, Steven C.
;   APPLICANT: Wiegand, Roger C.
;   TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;   FILE REFERENCE: 38-10(15849)B
;   CURRENT APPLICATION NUMBER: US/09/902,540
;   CURRENT FILING DATE: 2001-07-10
;   PRIOR APPLICATION NUMBER: 60/217,883
;   PRIOR FILING DATE: 2000-07-10
;   NUMBER OF SEQ ID NOS: 16825
;  SEQ ID NO 868
;   LENGTH: 7874
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-868

Alignment Scores:
Pred. No.:                         402      Length:          7874
Score:                            9.00      Matches:          9
Percent Similarity:             100.0%      Conservative:     0
Best Local Similarity:          100.0%      Mismatches:       0
Query Match:                      4.2%      Indels:           0
DB:                                  3      Gaps:             0

US-09-989-890-238 (1-212) x US-09-902-540-868 (1-7874)

Qy         56 SerAlaSerAlaAlaGlyIleAla 64
              ||||||||||||||||||||||||
Db       1546 AGTGCCAGCGCCGGCTGCGGGAATCGCT 1520


RESULT 19
US-09-902-540-966/c
;  Sequence 966, Application US/09902540
;  Patent No. 6833447
;  GENERAL INFORMATION:
;   APPLICANT: Goldman, Barry S.
;   APPLICANT: Hinkle, Gregory J.
;   APPLICANT: Slater, Steven C.
;   APPLICANT: Wiegand, Roger C.
;   TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;   FILE REFERENCE: 38-10(15849)B
;   CURRENT APPLICATION NUMBER: US/09/902,540
;   CURRENT FILING DATE: 2001-07-10
;   PRIOR APPLICATION NUMBER: 60/217,883
;   PRIOR FILING DATE: 2000-07-10
;   NUMBER OF SEQ ID NOS: 16825
;  SEQ ID NO 966
```

;   LENGTH: 9198
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-966

Alignment Scores:
Pred. No.:   463   Length: 9198
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-902-540-966 (1-9198)

Qy   50 SerGlyAlaGlyLeuProSerAlaSer 58
               ||||||||||||||||||||||||||
Db   1342 TCAGGTGCTGGGCCTTCCTTCAGCGAGT 1316

RESULT 20
US-09-949-016-15117/c
;   Sequence 15117, Application US/09949016
;   Patent No. 6812339
;   GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 15117
;   LENGTH: 9825
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15117

Alignment Scores:
Pred. No.:   492   Length: 9825
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-949-016-15117 (1-9825)

Qy   140 ThrArgGlyProProProGlnProLeu 148
               ||||||||||||||||||||||||||
Db   7661 ACTCGAGGACCCCCCCCCCAACCCCTA 7635

RESULT 21
US-09-919-172-97/c
;   Sequence 97, Application US/09919172
;   Patent No. 6673545
;   GENERAL INFORMATION:
;   APPLICANT: Faris, Mary
;   APPLICANT: Turner, Christopher M.
;   TITLE OF INVENTION: PROSTATE CANCER MARKERS
;   FILE REFERENCE: PA-0036 US
;   CURRENT APPLICATION NUMBER: US/09/919,172
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: 60/222,469
;   PRIOR FILING DATE: 2000-07-28
;   NUMBER OF SEQ ID NOS: 102
;   SOFTWARE: PERL Program

;   SEQ ID NO 97
;   LENGTH: 10432
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673545 2700132CB1
US-09-919-172-97

Alignment Scores:
Pred. No.:   519   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-919-172-97 (1-10432)

Qy   29 AlaLeuSerArgProLeuSerPro 37
               ||||||||||||||||||||||||
Db   5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 22
US-09-976-594-21/c
;   Sequence 21, Application US/09976594
;   Patent No. 6673549
;   GENERAL INFORMATION:
;   APPLICANT: Furness, Michael
;   APPLICANT: Buchbinder, Jenny
;   TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
;   FILE REFERENCE: PA-0041 US
;   CURRENT APPLICATION NUMBER: US/09/976,594
;   CURRENT FILING DATE: 2001-10-12
;   PRIOR APPLICATION NUMBER: 60/240,409
;   PRIOR FILING DATE: 2000-10-12
;   NUMBER OF SEQ ID NOS: 1143
;   SOFTWARE: PERL Program
;   SEQ ID NO 21
;   LENGTH: 10432
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673549 2700132CB1
US-09-976-594-21

Alignment Scores:
Pred. No.:   519   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-976-594-21 (1-10432)

Qy   29 AlaLeuLeuSerArgProLeuSerPro 37
               |||||||||||||||||||||||||
Db   5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 23
US-09-919-039-20/c
;   Sequence 20, Application US/09919039
;   Patent No. 6727066
;   GENERAL INFORMATION:
;   APPLICANT: Kaser, Matthew R.
;   TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
;   FILE REFERENCE: PA-0035 US
;   CURRENT APPLICATION NUMBER: US/09/919,039
;   CURRENT FILING DATE: 2002-09-09
;   PRIOR APPLICATION NUMBER: 60/222,113
;   PRIOR FILING DATE: 2000-07-28

```
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 20
; LENGTH: 10432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6727066 2700132CB1
US-09-919-039-20

Alignment Scores:
Pred. No.:              519           Length:         10432
Score:                  9.00          Matches:        9
Percent Similarity:     100.0%        Conservative:   0
Best Local Similarity:  100.0%        Mismatches:     0
Query Match:            4.2%          Indels:         0
DB:                     3             Gaps:           0

US-09-989-890-238 (1-212) x US-09-919-039-20 (1-10432)

Qy    29 AlaLeuSerArgProLeuSerPro 37
          ||||||||||||||||||||||||
Db  5063 GCTCTCTTTCACGCCACTTCCCCA 5037

RESULT 24
US-09-949-016-12988/c
; Sequence 12988, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12988
; LENGTH: 16013
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12988

Alignment Scores:
Pred. No.:              768           Length:         16013
Score:                  9.00          Matches:        9
Percent Similarity:     100.0%        Conservative:   0
Best Local Similarity:  100.0%        Mismatches:     0
Query Match:            4.2%          Indels:         0
DB:                     3             Gaps:           0

US-09-989-890-238 (1-212) x US-09-949-016-12988 (1-16013)

Qy   140 ThrArgGlyProProGlnProLeu 148
          ||||||||||||||||||||||||
Db  1691 ACTCGAGGACCCCCCCAACCCCTA 1665

RESULT 25
US-09-949-016-12410/c
; Sequence 12410, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12410
; LENGTH: 39498
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12410

Alignment Scores:
Pred. No.:              1.75e+03      Length:         39498
Score:                  9.00          Matches:        9
Percent Similarity:     100.0%        Conservative:   0
Best Local Similarity:  100.0%        Mismatches:     0
Query Match:            4.2%          Indels:         0
DB:                     3             Gaps:           0

US-09-989-890-238 (1-212) x US-09-949-016-12410 (1-39498)

Qy    25 ProSerLeuAlaLeuLeuSerArg 33
          |||||||||||||||||||||||||
Db 30929 CCGAGCTCCCTGGCTCTACTATCCAGG 30955

RESULT 26
US-09-949-016-16505/c
; Sequence 16505, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16505
; LENGTH: 39498
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16505

Alignment Scores:
Pred. No.:              1.75e+03      Length:         39498
Score:                  9.00          Matches:        9
Percent Similarity:     100.0%        Conservative:   0
Best Local Similarity:  100.0%        Mismatches:     0
Query Match:            4.2%          Indels:         0
DB:                     3             Gaps:           0

US-09-989-890-238 (1-212) x US-09-949-016-16505 (1-39498)

Qy    25 ProSerLeuAlaLeuLeuSerArg 33
          |||||||||||||||||||||||||
Db 30955 CCGAGCTCCCTGGCTCTACTATCCAGG 30929

RESULT 27
US-09-949-016-16348
; Sequence 16348, Application US/09949016
; Patent No. 6812339
```

```
;  GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 16348
;     LENGTH: 51252
;     TYPE: DNA
;     ORGANISM: Human
US-09-949-016-16348

Alignment Scores:
Pred. No.:              2.22e+03        Length:         51252
Score:                  9.00            Matches:        9
Percent Similarity:     100.0%          Conservative:   0
Best Local Similarity:  100.0%          Mismatches:     0
Query Match:            4.2%            Indels:         0
DB:                     3               Gaps:           0

US-09-989-890-238 (1-212) x US-09-949-016-16348 (1-51252)

Qy    24 LeuProSerSerLeuAlaLeuLeuSer 32
         |||||||||||||||||||||||||||
Db  1283 CTCCCAAGTTCTCTTGCCTTGCTATCC 1309

RESULT 28
US-09-949-016-15318
;  Sequence 15318, Application US/09949016
;  Patent No. 6812339
;  GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 15318
;     LENGTH: 79595
;     TYPE: DNA
;     ORGANISM: Human
US-09-949-016-15318

Alignment Scores:
Pred. No.:              3.32e+03        Length:         79595
Score:                  9.00            Matches:        9
Percent Similarity:     100.0%          Conservative:   0
Best Local Similarity:  100.0%          Mismatches:     0
Query Match:            4.2%            Indels:         0
DB:                     3               Gaps:           0

US-09-989-890-238 (1-212) x US-09-949-016-15318 (1-79595)

Qy     28 LeuAlaLeuLeuSerArgProLeuSer 36
          |||||||||||||||||||||||||||
Db  35245 TTAGCCCTTCTCTCTCGGCCTTTGTCT 35271


RESULT 29
US-09-949-016-15642/c
;  Sequence 15642, Application US/09949016
;  Patent No. 6812339
;  GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 15642
;     LENGTH: 79634
;     TYPE: DNA
;     ORGANISM: Human
US-09-949-016-15642

Alignment Scores:
Pred. No.:              3.32e+03        Length:         79634
Score:                  9.00            Matches:        9
Percent Similarity:     100.0%          Conservative:   0
Best Local Similarity:  100.0%          Mismatches:     0
Query Match:            4.2%            Indels:         0
DB:                     3               Gaps:           0

US-09-989-890-238 (1-212) x US-09-949-016-15642 (1-79634)

Qy     22 ArgAlaLeuProSerSerLeuAlaLeu 30
          |||||||||||||||||||||||||||
Db  56785 AGAGCCCTTCCATCCTCCTTAGCCCTC 56759


RESULT 30
US-09-949-016-12345/c
;  Sequence 12345, Application US/09949016
;  Patent No. 6812339
;  GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 12345
;     LENGTH: 86414
;     TYPE: DNA
;     ORGANISM: Human
;     FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: (1)...(86414)
;     OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12345

Alignment Scores:
Pred. No.:              3.58e+03        Length:         86414
Score:                  9.00            Matches:        9
Percent Similarity:     100.0%          Conservative:   0
```

```
Best Local Similarity:  100.0%      Mismatches:   0
Query Match:            4.2%        Indels:       0
DB:                                Gaps:         0

US-09-989-890-238 (1-212) x US-09-949-016-12345 (1-86414)

Qy    31 LeuSerArgProLeuSerProPro 39
         ||||||||||||||||||||||||
Db  70274 CTTTCCAGACCACTGTCCCCTCCCCCCA 70248

RESULT 31
US-09-949-016-15758/c
; Sequence 15758, Application US/09994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 15758
; LENGTH: 86414
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(86414)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15758

Alignment Scores:
Pred. No.:              3.58e+03      Length:        86414
Score:                  9.00         Matches:       9
Percent Similarity:     100.0%       Conservative:  0
Best Local Similarity:  100.0%       Mismatches:    0
Query Match:            4.2%         Indels:        0
DB:                                 Gaps:          0

US-09-989-890-238 (1-212) x US-09-949-016-15758 (1-86414)

Qy    31 LeuSerArgProLeuSerProPro 39
         ||||||||||||||||||||||||
Db  70274 CTTTCCAGACCACTGTCCCCTCCCCCCA 70248

RESULT 32
US-09-949-016-12751
; Sequence 12751, Application US/09994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12751


; LENGTH: 148794
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12751

Alignment Scores:
Pred. No.:              5.88e+03      Length:        148794
Score:                  9.00         Matches:       9
Percent Similarity:     100.0%       Conservative:  0
Best Local Similarity:  100.0%       Mismatches:    0
Query Match:            4.2%         Indels:        0
DB:                                 Gaps:          0

US-09-989-890-238 (1-212) x US-09-949-016-12751 (1-148794)

Qy    28 LeuAlaLeuLeuSerArgProLeuSer 36
         ||||||||||||||||||||||||||
Db  56444 TTAGCCCTTCTCTCTCGGCCCTTGTCT 56470

RESULT 33
US-09-949-016-14157/c
; Sequence 14157, Application US/09994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 14157
; LENGTH: 455726
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(455726)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14157

Alignment Scores:
Pred. No.:              1.63e+04      Length:        455726
Score:                  9.00         Matches:       9
Percent Similarity:     100.0%       Conservative:  0
Best Local Similarity:  100.0%       Mismatches:    0
Query Match:            4.2%         Indels:        0
DB:                                 Gaps:          0

US-09-989-890-238 (1-212) x US-09-949-016-14157 (1-455726)

Qy   127 ThrProMetLeuSerSerLysAlaSer 135
         ||||||||||||||||||||||||||
Db 189887 ACACCTATGCTCTCCTCCAAAGCTTCA 189861

RESULT 34
US-09-949-016-11940/c
; Sequence 11940, Application US/09994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11940
; LENGTH: 481115
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(481115)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11940

Alignment Scores:
Pred. No.:   1.71e+04   Length: 481115
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-949-016-11940 (1-481115)

Qy   127 ThrProMetLeuSerSerLysAlaSer 135
            |||||||||||||||||||||||||||
Db   167276 ACACCTATGCTCTCCTCCAAAGCTTCA 167250

RESULT 35
US-10-131-827-69
; Sequence 69, Application US/10131827
; Patent No. 6905827
; GENERAL INFORMATION:
; APPLICANT: Wohlgemuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 506612000120
; CURRENT APPLICATION NUMBER: US/10/131,827
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69
; LENGTH: 50
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-69

Alignment Scores:
Pred. No.:   32.2   Length: 50
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-10-131-827-69 (1-50)

Qy   25 ProSerLeuAlaLeuLeuSer 32
        ||||||||||||||||||||||
Db   23 CCTTCAAGCCTAGCCCTTCTCTCA 46

, RESULT 36

US-10-131-827-4540
; Sequence 4540, Application US/10131827
; Patent No. 6905827
; GENERAL INFORMATION:
; APPLICANT: Wohlgemuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 506612000120
; CURRENT APPLICATION NUMBER: US/10/131,827
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4540
; LENGTH: 50
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-4540

Alignment Scores:
Pred. No.:   32.2   Length: 50
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-10-131-827-4540 (1-50)

Qy   25 ProSerLeuAlaLeuLeuSer 32
        ||||||||||||||||||||||
Db   4 CCTTCAAGCCTAGCCCTTCTCTCA 27

RESULT 37
US-08-474-633A-96
; Sequence 96, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; APPLICANT: COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
;

```
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  302-992-4931
;     TELEFAX:  302-773-0164
;     TELEX:  835420
;   INFORMATION FOR SEQ ID NO:  96:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  55 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
US-08-474-633A-96

Alignment Scores:
Pred. No.:
Score:                     35.1      Length:          55
Percent Similarity:       100.0%     Matches:          8
Best Local Similarity:    100.0%     Conservative:     0
Query Match:                3.8%     Mismatches:       0
DB:                                  Indels:           0
                                     Gaps:             0

US-09-989-890-238 (1-212) x US-08-474-633A-96 (1-55)

Qy   33 ArgProLeuSerProProAla  40
        ||||||||||||||||||||||||
Db   11 CGCCCCTTAAGTCCACCGCCAGCC  34


RESULT 38
US-08-474-633A-97/C
; Sequence 97, Application US/08474633A
; Patent No. 5773691
;   GENERAL INFORMATION:
;     APPLICANT:  E. I. DU PONT DE NEMOURS AND
;     APPLICANT:  COMPANY
;     TITLE OF INVENTION:  CHIMERIC GENES AND
;     TITLE OF INVENTION:  METHODS FOR INCREASING
;     TITLE OF INVENTION:  INCREASING THE LYSINE
;     TITLE OF INVENTION:  AND THREONINE CONTENT
;     TITLE OF INVENTION:  OF THE SEEDS OF PLANTS
;   NUMBER OF SEQUENCES:  107
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  E. I. DU PONT DE NEMOURS
;     ADDRESSEE:  AND COMPANY
;     STREET:  1007 MARKET STREET
;     CITY:  WILMINGTON
;     STATE:  DELAWARE
;     COUNTRY:  U.S.A.
;     ZIP:  19898
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  FLOPPY DISK
;     COMPUTER:  IBM PC COMPATIBLE
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  MICROSOFT WORD VERSION 2.0C
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/474,633A
;     FILING DATE:
;     CLASSIFICATION:  800
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  BARBARA C. SIEGELL
;     REGISTRATION NUMBER:  30,684
;     REFERENCE/DOCKET NUMBER:  BB-1037-C
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  302-992-4931
;     TELEFAX:  302-773-0164
;     TELEX:  835420
;   INFORMATION FOR SEQ ID NO:  97:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  55 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
US-08-474-633A-97
```

```
Alignment Scores:
Pred. No.:
Score:                     35.1      Length:          55
Percent Similarity:       100.0%     Matches:          8
Best Local Similarity:    100.0%     Conservative:     0
Query Match:                3.8%     Mismatches:       0
DB:                                  Indels:           0
                                     Gaps:             0

US-09-989-890-238 (1-212) x US-08-474-633A-97 (1-55)

Qy   33 ArgProLeuSerProProAla  40
        ||||||||||||||||||||||||
Db   49 CGCCCCTTAAGTCCACCGCCAGCC  26


RESULT 39
US-08-737-524B-20
; Sequence 20, Application US/08737524B
; Patent No. 5912414
;   GENERAL INFORMATION:
;     APPLICANT:  CARL SAVERIO FALCO
;     APPLICANT:  DOMINICK ANTHONY GUIDA, JR.
;     APPLICANT:  MARY ELIZABETH HARNETT LOCKE
;     TITLE OF INVENTION:  NUCLEIC ACID FRAGMENTS, CHIMERIC
;     TITLE OF INVENTION:  GENES AND METHODS FOR INCREASING
;     TITLE OF INVENTION:  THE METHIONINE CONTENT OF THE SEEDS
;     TITLE OF INVENTION:  OF PLANTS
;   NUMBER OF SEQUENCES:  27
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;     STREET:  1007 MARKET STREET
;     CITY:  WILMINGTON
;     STATE:  DELAWARE
;     COUNTRY:  UNITED STATES OF AMERICA
;     ZIP:  19898
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  DISKETTE, 3.50 INCH
;     COMPUTER:  IBM PC COMPATIBLE
;     OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;     SOFTWARE:  MICROSOFT WORD FOR WINDOWS 95 (7.0)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/737,524B
;     FILING DATE:
;     CLASSIFICATION:  800
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  LYNNE M. CHRISTENBURY
;     REGISTRATION NUMBER:  30,971
;     REFERENCE/DOCKET NUMBER:  BB-1059-A
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  302-992-5481
;     TELEFAX:  302-773-0164
;     TELEX:  835420
;   INFORMATION FOR SEQ ID NO:  20:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  55 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
US-08-737-524B-20

Alignment Scores:
Pred. No.:
Score:                     35.1      Length:          55
Percent Similarity:       100.0%     Matches:          8
Best Local Similarity:    100.0%     Conservative:     0
Query Match:                3.8%     Mismatches:       0
DB:                                  Indels:           0
                                     Gaps:             0

US-09-989-890-238 (1-212) x US-08-737-524B-20 (1-55)

Qy   33 ArgProLeuSerProProAla  40
        ||||||||||||||||||||||||
```

```
Db            11  CGCCCCTTAAGTCCACCGCCAGCC  34

RESULT 40
US-08-737-524B-21/c
; Sequence 21, Application US/08737524B
; Patent No. 5912414
; GENERAL INFORMATION:
;  APPLICANT:  CARL SAVERIO FALCO
;  APPLICANT:  DOMINICK ANTHONY GUIDA, JR.
;  APPLICANT:  MARY ELIZABETH HARNETT LOCKE
;  TITLE OF INVENTION:  NUCLEIC ACID FRAGMENTS, CHIMERIC
;  TITLE OF INVENTION:  GENES AND METHODS FOR INCREASING
;  TITLE OF INVENTION:  THE METHIONINE CONTENT OF THE SEEDS
;  TITLE OF INVENTION:  OF PLANTS
;  NUMBER OF SEQUENCES: 27
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;   STREET: 1007 MARKET STREET
;   CITY: WILMINGTON
;   STATE: DELAWARE
;   COUNTRY: UNITED STATES OF AMERICA
;   ZIP: 19898
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: DISKETTE, 3.50 INCH
;   COMPUTER: IBM PC COMPATIBLE
;   OPERATING SYSTEM: MICROSOFT WINDOWS 95
;   SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (7.0)
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/737,524B
;   FILING DATE:
;   CLASSIFICATION:  800
;  ATTORNEY/AGENT INFORMATION:
;   NAME: LYNNE M. CHRISTENBURY
;   REGISTRATION NUMBER: 30,971
;   REFERENCE/DOCKET NUMBER: BB-1059-A
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 302-992-5481
;   TELEFAX: 302-773-0164
;   TELEX: 835420
;  INFORMATION FOR SEQ ID NO: 21:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 55 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-737-524B-21

Alignment Scores:
Pred. No.:                             35.1        Length:          55
Score:                                 8.00        Matches:          8
Percent Similarity:                   100.0%       Conservative:     0
Best Local Similarity:                100.0%       Mismatches:       0
Query Match:                            3.8%       Indels:           0
DB:                                    2           Gaps:             0

US-09-989-890-238 (1-212) x US-08-737-524B-21 (1-55)

Qy         33  ArgProLeuSerProProProAla  40
               ||||||||||||||||||||||||
Db         49  CGCCCCTTAAGTCCACCGCCAGCC  26

RESULT 41
US-08-823-771-96
; Sequence 96, Application US/08823771
; Patent No. 6459019
; GENERAL INFORMATION:
;  APPLICANT: E. I. DU PONT DE NEMOURS AND
;   COMPANY
;  TITLE OF INVENTION: CHIMERIC GENES AND
;   METHODS FOR INCREASING
;   INCREASING THE LYSINE
;   AND THREONINE CONTENT
```

```
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
 ADDRESSEE: E. I. DU PONT DE NEMOURS
  AND COMPANY
 STREET: 1007 MARKET STREET
 CITY: WILMINGTON
 STATE: DELAWARE
 COUNTRY: U.S.A.
 ZIP: 19898
COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/823,771
 FILING DATE: 24-Mar-1997
 CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/474,633
 FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
 NAME: BARBARA C. SIEGELL
 REGISTRATION NUMBER: 30,684
 REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
 TELEPHONE: 302-992-4931
 TELEFAX: 302-773-0164
 TELEX: 835420
INFORMATION FOR SEQ ID NO: 96:
 SEQUENCE CHARACTERISTICS:
  LENGTH: 55 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 96:
US-08-823-771-96

Alignment Scores:
Pred. No.:                             35.1        Length:          55
Score:                                 8.00        Matches:          8
Percent Similarity:                   100.0%       Conservative:     0
Best Local Similarity:                100.0%       Mismatches:       0
Query Match:                            3.8%       Indels:           0
DB:                                    3           Gaps:             0

US-09-989-890-238 (1-212) x US-08-823-771-96 (1-55)

Qy         33  ArgProLeuSerProProProAla  40
               ||||||||||||||||||||||||
Db         11  CGCCCCTTAAGTCCACCGCCAGCC  34

RESULT 42
US-08-823-771-97/c
; Sequence 97, Application US/08823771
; Patent No. 6459019
; GENERAL INFORMATION:
;  APPLICANT: E. I. DU PONT DE NEMOURS AND
;   COMPANY
;  TITLE OF INVENTION: CHIMERIC GENES AND
;   METHODS FOR INCREASING
;   INCREASING THE LYSINE
;   AND THREONINE CONTENT
;  NUMBER OF SEQUENCES: 107
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: E. I. DU PONT DE NEMOURS
;    AND COMPANY
;   STREET: 1007 MARKET STREET
;   CITY: WILMINGTON
;   STATE: DELAWARE
;   COUNTRY: U.S.A.
;   ZIP: 19898
```

COMPUTER READABLE FORM:
  MEDIUM TYPE: FLOPPY DISK
  COMPUTER: IBM PC COMPATIBLE
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/823,771
  FILING DATE: 24-Mar-1997
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/474,633
  FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
  NAME: BARBARA C. SIEGELL
  REGISTRATION NUMBER: 30,684
  REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 302-992-4931
  TELEFAX: 302-773-0164
  TELEX: 835420
INFORMATION FOR SEQ ID NO: 97:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 55 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  SEQUENCE DESCRIPTION: SEQ ID NO: 97:
US-08-823-771-97

Alignment Scores:
Pred. No.:   35.1   Length: 55
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-08-823-771-97 (1-55)

Qy   33 ArgProLeuSerProProProAla 40
         ||||||||||||||||||||||||
Db   49 CGCCCCTTAAGTCCACCGCCAGCC 26

RESULT 43
PCT-US93-02480-19
; Sequence 19, Application PC/TUS9302480
; GENERAL INFORMATION:
  APPLICANT: E.I. DU PONT DE NEMOURS AND
  APPLICANT: COMPANY
  TITLE OF INVENTION: NUCLEIC ACID
  TITLE OF INVENTION: FRAGMENTS AND
  TITLE OF INVENTION: METHODS FOR
  TITLE OF INVENTION: INCREASING THE
  TITLE OF INVENTION: LYSINE AND THREONINE
  TITLE OF INVENTION: CONTENT OF THE SEEDS
  TITLE OF INVENTION: OF PLANTS
  NUMBER OF SEQUENCES: 24
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: E.I. DU PONT DE NEMOURS
    ADDRESSEE: AND COMPANY
    STREET: 1007 MARKET STREET
    CITY: WILMINGTON
    STATE: DELAWARE
    COUNTRY: U.S.A.
    ZIP: 19898
  COMPUTER READABLE FORM:
    MEDIUM TYPE: DISKETTE, 3.50 INCH
    COMPUTER: MACINTOSH
    OPERATING SYSTEM: MACINTOSH, 6.0
    SOFTWARE: MICROSOFT WORD, 4.0
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: PCT/US93/02480

  FILING DATE: 19930318
  CLASSIFICATION:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 07/855,414
  FILING DATE: 19 MARCH 1992
ATTORNEY/AGENT INFORMATION:
  NAME: LINDA AXAMETHY FLOYD
  REGISTRATION NUMBER: 33,692
  REFERENCE/DOCKET NUMBER: BB-1037-A
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 302-992-4929
  TELEFAX: 302-892-7949
  TELEX: 835420
INFORMATION FOR SEQ ID NO: 19:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 55 base pairs
    TYPE: NUCLEIC ACID
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
PCT-US93-02480-19

Alignment Scores:
Pred. No.:   35.1   Length: 55
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x PCT-US93-02480-19 (1-55)

Qy   33 ArgProLeuSerProProProAla 40
         ||||||||||||||||||||||||
Db   11 CGCCCCTTAAGTCCACCGCCAGCC 34

RESULT 44
PCT-US93-02480-20/c
; Sequence 20, Application PC/TUS9302480
; GENERAL INFORMATION:
  APPLICANT: E.I. DU PONT DE NEMOURS AND
  APPLICANT: COMPANY
  TITLE OF INVENTION: NUCLEIC ACID
  TITLE OF INVENTION: FRAGMENTS AND
  TITLE OF INVENTION: METHODS FOR
  TITLE OF INVENTION: INCREASING THE
  TITLE OF INVENTION: LYSINE AND THREONINE
  TITLE OF INVENTION: CONTENT OF THE SEEDS
  TITLE OF INVENTION: OF PLANTS
  NUMBER OF SEQUENCES: 24
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: E.I. DU PONT DE NEMOURS
    ADDRESSEE: AND COMPANY
    STREET: 1007 MARKET STREET
    CITY: WILMINGTON
    STATE: DELAWARE
    COUNTRY: U.S.A.
    ZIP: 19898
  COMPUTER READABLE FORM:
    MEDIUM TYPE: DISKETTE, 3.50 INCH
    COMPUTER: MACINTOSH
    OPERATING SYSTEM: MACINTOSH, 6.0
    SOFTWARE: MICROSOFT WORD, 4.0
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: PCT/US93/02480
    FILING DATE: 19930318
    CLASSIFICATION:
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/855,414
    FILING DATE: 19 MARCH 1992
  ATTORNEY/AGENT INFORMATION:
    NAME: LINDA AXAMETHY FLOYD
    REGISTRATION NUMBER: 33,692

Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x PCT-US95-08501-5 (1-55)

Qy   33 ArgProLeuSerProProAla 40
           |||||||||||||||||||||
Db   11 CGCCCCTTAAGTCCACCGCCAGCC 34

RESULT 46
PCT-US95-08501-6/c
; Sequence 6, Application PC/TUS9508501
  GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHOD FOR INCREASING THE
; TITLE OF INVENTION: THREONINE CONTENT OF
; TITLE OF INVENTION: THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08501
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1063-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-08501-6

Alignment Scores:
Pred. No.:   35.1   Length: 55
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x PCT-US95-08501-6 (1-55)

Qy   33 ArgProLeuSerProProAla 40
           |||||||||||||||||||||
Db   49 CGCCCCTTAAGTCCACCGCCAGCC 26

RESULT 47
US-09-513-999C-12253/c
; Sequence 12253, Application US/09513999C
; Patent No. 6783961

; REFERENCE/DOCKET NUMBER: BB-1037-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4929
; TELEFAX: 302-892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-02480-20

Alignment Scores:
Pred. No.:   35.1   Length: 55
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x PCT-US93-02480-20 (1-55)

Qy   33 ArgProLeuSerProProAla 40
           |||||||||||||||||||||
Db   49 CGCCCCTTAAGTCCACCGCCAGCC 26

RESULT 45
PCT-US95-08501-5
; Sequence 5, Application PC/TUS9508501
  GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHOD FOR INCREASING THE
; TITLE OF INVENTION: THREONINE CONTENT OF
; TITLE OF INVENTION: THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08501
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1063-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-08501-5

Alignment Scores:
Pred. No.:   35.1   Length: 55

; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 12253
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 153
; OTHER INFORMATION: r=a or g
US-09-513-999C-12253

Alignment Scores:
Pred. No.:   104
Score: 8.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 3.8%
DB: 3

Length: 180
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x US-09-513-999C-12253 (1-180)

Qy   31 LeuSerArgProLeuSerProPro 38
         ||||||||||||||||||||||||
Db   83 CTCTCTCGCCCTCTCTCCCCCCCA 60

RESULT 48
US-09-949-016-807
; Sequence 807, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 807
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-807

Alignment Scores:
Pred. No.:   150
Score: 8.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 3.8%
DB: 3

Length: 270
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x US-09-949-016-807 (1-270)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62

Db   39 CCCTCAGCCTTCAGCAGCAGCAGGT 62

RESULT 49
US-09-949-016-2217
; Sequence 2217, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2217
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2217

Alignment Scores:
Pred. No.:   183
Score: 8.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 3.8%
DB: 3

Length: 336
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x US-09-949-016-2217 (1-336)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62

Db   51 CCCTCAGCCTTCAGCAGCAGCAGGT 74

RESULT 50
US-09-123-912-97
; Sequence 97, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C1
; CURRENT APPLICATION NUMBER: US/09/123,912A
; CURRENT FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/040,802
; PRIOR FILING DATE: 1998-03-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 97
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (12)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (308)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-97

Alignment Scores:
Pred. No.:   235
Score: 8.00

Length: 441
Matches: 8

```
Percent Similarity:        100.0%       Conservative:         0
Best Local Similarity:     100.0%       Mismatches:           0
Query Match:                 3.8%       Indels:               0
DB:                            3        Gaps:                 0

US-09-989-890-238 (1-212) x US-09-123-912-97 (1-441)

Qy     55 ProSerAlaSerAlaAlaGly 62
          ||||||||||||||||||||
Db     92 CCCTCAGCTTCAGCAGCAGGT 115

RESULT 51
US-09-643-597-97
; Sequence 97, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
   APPLICANT: Wang, Tongtong
   APPLICANT: Fan, Liqun
   APPLICANT: Kalos, Michael D.
   APPLICANT: Bangur, Chaitanya S.
   APPLICANT: Hosken, Nancy
   APPLICANT: Fanger, Gary R.
   APPLICANT: Li, Samuel X.
   APPLICANT: Wang, Aijun
   APPLICANT: Skeiky, Yasir A.W.
   APPLICANT: Henderson, Robert A.
   APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 97
;   LENGTH: 441
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(441)
;   OTHER INFORMATION: n = A,T,C or G
US-09-643-597-97

Alignment Scores:
Pred. No.:                   235        Length:             441
Score:                      8.00        Matches:              8
Percent Similarity:        100.0%       Conservative:         0
Best Local Similarity:     100.0%       Mismatches:           0
Query Match:                 3.8%       Indels:               0
DB:                            3        Gaps:                 0

US-09-989-890-238 (1-212) x US-09-643-597-97 (1-441)

Qy     55 ProSerAlaSerAlaAlaGly 62
          ||||||||||||||||||||
Db     92 CCCTCAGCTTCAGCAGCAGGT 115

RESULT 52
US-09-480-884A-97
; Sequence 97, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
   APPLICANT: Wang, Tongtong
   APPLICANT: Fan, Liqun
   APPLICANT: Hosken, Nancy A.
   APPLICANT: Kalos, Michael D.
   APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C6
; CURRENT APPLICATION NUMBER: US/09/480,884A
; CURRENT FILING DATE: 2001-08-27
```

```
; NUMBER OF SEQ ID NOS: 330
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 97
;   LENGTH: 441
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(441)
;   OTHER INFORMATION: n = A,T,C or G
US-09-480-884A-97

Alignment Scores:
Pred. No.:                   235        Length:             441
Score:                      8.00        Matches:              8
Percent Similarity:        100.0%       Conservative:         0
Best Local Similarity:     100.0%       Mismatches:           0
Query Match:                 3.8%       Indels:               0
DB:                            3        Gaps:                 0

US-09-989-890-238 (1-212) x US-09-480-884A-97 (1-441)

Qy     55 ProSerAlaSerAlaAlaGly 62
          ||||||||||||||||||||
Db     92 CCCTCAGCTTCAGCAGCAGGT 115

RESULT 53
US-09-542-615A-97
; Sequence 97, Application US/09542615A
; Patent No. 6518256
; GENERAL INFORMATION:
   APPLICANT: Wang, Tongtong
   APPLICANT: Fan, Liqun
   APPLICANT: Kalos, Michael D.
   APPLICANT: Bangur, Chaitanya S.
   APPLICANT: Hosken, Nancy A.
   APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C8
; CURRENT APPLICATION NUMBER: US/09/542,615A
; CURRENT FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 350
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 97
;   LENGTH: 441
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(441)
;   OTHER INFORMATION: n = A,T,C or G
US-09-542-615A-97

Alignment Scores:
Pred. No.:                   235        Length:             441
Score:                      8.00        Matches:              8
Percent Similarity:        100.0%       Conservative:         0
Best Local Similarity:     100.0%       Mismatches:           0
Query Match:                 3.8%       Indels:               0
DB:                            3        Gaps:                 0

US-09-989-890-238 (1-212) x US-09-542-615A-97 (1-441)

Qy     55 ProSerAlaSerAlaAlaGly 62
          ||||||||||||||||||||
Db     92 CCCTCAGCTTCAGCAGCAGGT 115

RESULT 54
US-09-606-421B-97
; Sequence 97, Application US/09606421B
; Patent No. 6531315
; GENERAL INFORMATION:
```

; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C9
; CURRENT APPLICATION NUMBER: US/09/606,421B
; CURRENT FILING DATE: 2000-06-28
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 97
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(441)
; OTHER INFORMATION: n = A,T,C or G
US-09-606-421B-97

Alignment Scores:
Pred. No.:   235
Score: 8.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 3.8%
DB: 3

US-09-989-890-238 (1-212) x US-09-606-421B-97 (1-441)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
          ||||||||||||||||||||||||
Db   92 CCCTCAGCTTCAGCAGCAGCAGGT 115

RESULT 55
US-09-221-107-97
; Sequence 97, Application US/09221107
; Patent No. 6660838
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C2
; CURRENT APPLICATION NUMBER: US/09/221,107
; CURRENT FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 97
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (12)
; OTHER INFORMATION: Where n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (308)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-221-107-97

Alignment Scores:
Pred. No.:   235
Score: 8.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 3.8%
DB: 3

Length: 441
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

Length: 441
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x US-09-221-107-97 (1-441)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
          ||||||||||||||||||||||||
Db   92 CCCTCAGCTTCAGCAGCAGCAGGT 115

RESULT 56
US-09-466-396A-97
; Sequence 97, Application US/09466396A
; Patent No. 6696247
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C4
; CURRENT APPLICATION NUMBER: US/09/466,396A
; CURRENT FILING DATE: 1999-12-17
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 97
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(441)
; OTHER INFORMATION: n = A,T,C or G
US-09-466-396A-97

Alignment Scores:
Pred. No.:   235
Score: 8.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 3.8%
DB: 3

Length: 441
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x US-09-466-396A-97 (1-441)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
          ||||||||||||||||||||||||
Db   92 CCCTCAGCTTCAGCAGCAGCAGGT 115

RESULT 57
US-09-476-496A-97
; Sequence 97, Application US/09476496A
; Patent No. 6706262
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF
; TITLE OF INVENTION: LUNG CANCER
; FILE REFERENCE: 210121.455C5
; CURRENT APPLICATION NUMBER: US/09/476,496A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 97
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(441)
; OTHER INFORMATION: n = A,T,C or G
US-09-476-496A-97

Alignment Scores:
Pred. No.:   235
Score: 8.00

Length: 441
Matches: 8

Fri Mar 17 09:43:20 2006

us-09-989-890-238.oligo_p2n.rni

```
Percent Similarity:          100.0%    Conservative:        0
Best Local Similarity:       100.0%    Mismatches:          0
Query Match:                   3.8%    Indels:              0
DB:                                    Gaps:                0

US-09-989-890-238 (1-212) x US-09-476-496A-97 (1-441)

Qy    55  ProSerAlaSerAlaAlaAlaGly 62
              |||||||||||||||||||||
Db    92  CCCTCAGCTTCAGCAGCAGCAGGT 115

RESULT 58
US-09-630-940B-97
; Sequence 97, Application US/09630940B
; Patent No. 6737514
; GENERAL INFORMATION:
  APPLICANT: Wang, Tongtong
  APPLICANT: Fan, Liqun
  APPLICANT: Kalos, Michael D.
  APPLICANT: Bangur, Chaitanya S.
  APPLICANT: Hosken, Nancy
  APPLICANT: Fanger, Gary R.
  APPLICANT: Li, Samuel X.
  APPLICANT: Wang, Aijun
  APPLICANT: Skeiky, Yasir A.W.
  APPLICANT: Henderson, Robert A.
  APPLICANT: McNeill, Patricia D.
  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
  TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
  FILE REFERENCE: 210121.455C10
  CURRENT APPLICATION NUMBER: US/09/630,940B
  CURRENT FILING DATE: 2000-08-02
  NUMBER OF SEQ ID NOS: 367
  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 97
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(441)
; OTHER INFORMATION: n = A,T,C or G
US-09-630-940B-97

Alignment Scores:
Pred. No.:                      235    Length:            441
Score:                         8.00    Matches:             8
Percent Similarity:          100.0%    Conservative:        0
Best Local Similarity:       100.0%    Mismatches:          0
Query Match:                   3.8%    Indels:              0
DB:                               3    Gaps:                0

US-09-989-890-238 (1-212) x US-09-630-940B-97 (1-441)

Qy    55  ProSerAlaSerAlaAlaAlaGly 62
              |||||||||||||||||||||
Db    92  CCCTCAGCTTCAGCAGCAGCAGGT 115

RESULT 59
US-09-285-479-97
; Patent No. 6821518
; GENERAL INFORMATION:
  APPLICANT: Wang, Tongtong
  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF
  TITLE OF INVENTION: LUNG CANCER
  FILE REFERENCE: 210121.455C3
  CURRENT APPLICATION NUMBER: US/09/285,479
  CURRENT FILING DATE: 1999-04-02
  NUMBER OF SEQ ID NOS: 172
  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 97
; LENGTH: 441
```

```
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(441)
; OTHER INFORMATION: n = A,T,C or G
US-09-285-479-97

Alignment Scores:
Pred. No.:                      235    Length:            441
Score:                         8.00    Matches:             8
Percent Similarity:          100.0%    Conservative:        0
Best Local Similarity:       100.0%    Mismatches:          0
Query Match:                   3.8%    Indels:              0
DB:                               3    Gaps:                0

US-09-989-890-238 (1-212) x US-09-285-479-97 (1-441)

Qy    55  ProSerAlaSerAlaAlaAlaGly 62
              |||||||||||||||||||||
Db    92  CCCTCAGCTTCAGCAGCAGCAGGT 115

RESULT 60
US-10-007-700-97
; Sequence 97, Application US/10007700
; Patent No. 6960570
; GENERAL INFORMATION:
  APPLICANT: Wang, Tongtong
  APPLICANT: Wang, Aijun
  APPLICANT: Skeiky, Yasir A.W.
  APPLICANT: Li, Samual X.
  APPLICANT: Kalos, Michael D.
  APPLICANT: Henderson, Robert A.
  APPLICANT: McNeill, Patricia D.
  APPLICANT: Fanger, Neil
  APPLICANT: Retter, Marc W.
  APPLICANT: Durham, Margarita
  APPLICANT: Fanger, Gary R.
  APPLICANT: Vedvick, Thomas S.
  APPLICANT: Carter, Darrick
  APPLICANT: Watanabe, Yoshihiro
  APPLICANT: Peckman, David W.
  APPLICANT: Cai, Feng
  APPLICANT: Foy, Teresa M.
  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
  TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
  FILE REFERENCE: 210121.455C17
  CURRENT APPLICATION NUMBER: US/10/007,700
  CURRENT FILING DATE: 2001-11-30
  NUMBER OF SEQ ID NOS: 469
  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12, 308
; OTHER INFORMATION: n = A,T,C or G
US-10-007-700-97

Alignment Scores:
Pred. No.:                      235    Length:            441
Score:                         8.00    Matches:             8
Percent Similarity:          100.0%    Conservative:        0
Best Local Similarity:       100.0%    Mismatches:          0
Query Match:                   3.8%    Indels:              0
DB:                               3    Gaps:                0

US-09-989-890-238 (1-212) x US-10-007-700-97 (1-441)

Qy    55  ProSerAlaSerAlaAlaAlaGly 62
              |||||||||||||||||||||
```

Db   92 CCCTCAGCTTCAGCAGCAGCAGGT 115

RESULT 61
US-09-270-767-6252
; Sequence 6252, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;   APPLICANT: Homburger et al.
;   TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;   FILE REFERENCE: File Reference: 7326-094
;   CURRENT APPLICATION NUMBER: US/09/270,767
;   CURRENT FILING DATE: 1999-03-17
;   NUMBER OF SEQ ID NOS: 62517
; SEQ ID NO 6252
;   LENGTH: 479
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-6252

Alignment Scores:
Pred. No.:   253   Length: 479
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-270-767-6252 (1-479)

Qy   58 SerAlaAlaAlaGlyIleAlaSer 65
           ||||||||||||||||||||||||
Db   222 TCTGCTGCGCAGCAGGGATCGCATCT 245

RESULT 62
US-09-270-767-21534
; Sequence 21534, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;   APPLICANT: Homburger et al.
;   TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;   FILE REFERENCE: File Reference: 7326-094
;   CURRENT APPLICATION NUMBER: US/09/270,767
;   CURRENT FILING DATE: 1999-03-17
;   NUMBER OF SEQ ID NOS: 62517
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21534
;   LENGTH: 479
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-21534

Alignment Scores:
Pred. No.:   253   Length: 479
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-270-767-21534 (1-479)

Qy   58 SerAlaAlaAlaGlyIleAlaSer 65
           ||||||||||||||||||||||||
Db   222 TCTGCTGCGCAGCAGGGATCGCATCT 245

RESULT 63
US-08-913-362-29/c
; Sequence 29, Application US/08913362
; Patent No. 6287574
; GENERAL INFORMATION:
;   APPLICANT: Brodeur, Bernard R
;   APPLICANT: Martin, Denis

;   APPLICANT: Hamel, Josee
;   APPLICANT: Rioux, Clement
;   TITLE OF INVENTION: PROTEINASE K RESISTANT SURFACE PROTEIN
;   TITLE OF INVENTION: OF NEISSERIA MENINGITIDIS
;   NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 3000 K Street, N.W., Suite 500
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20007-5109
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/913,362
;   FILING DATE: 13-NOV-1997
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/406,362
;   FILING DATE: 17-MAR-1995
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/001,983
;   FILING DATE: 04-AUG-1995
; ATTORNEY/AGENT INFORMATION:
;   NAME: Bent, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 047998/0128
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202)672-5300
;   TELEFAX: (202)672-5399
;   TELEX: 904136
; INFORMATION FOR SEQ ID NO: 29:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 528 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: other nucleic acid
;   DESCRIPTION: /desc = "consensus"
US-08-913-362-29

Alignment Scores:
Pred. No.:   277   Length: 528
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-08-913-362-29 (1-528)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
           ||||||||||||||||||||||||
Db   62 CCTTCCGCCAGTGCGGCGGCCGGG 39

RESULT 64
US-09-949-002-191/c
; Sequence 191, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;   TITLE OF INVENTION: AND USES THEREOF
;   FILE REFERENCE: CL000790
;   CURRENT APPLICATION NUMBER: US/09/949,002
;   CURRENT FILING DATE: 2000-01-28
;   PRIOR APPLICATION NUMBER: 60/231,401
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 10823

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 191
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-191

Alignment Scores:
Pred. No.:                   296        Length:          569
Score:                       8.00       Matches:           8
Percent Similarity:          100.0%     Conservative:      0
Best Local Similarity:       100.0%     Mismatches:        0
Query Match:                 3.8%       Indels:            0
DB:                          3          Gaps:              0

US-09-989-890-238 (1-212) x US-08-991-789A-222 (1-578)

Qy    30  LeuLeuSerArgSerPro 37
          ||||||||||||||||||
Db    424 CTCCTCAGCAGGGCCCTTGAGTCCC 401

RESULT 65
US-08-991-789A-222
; Sequence 222, Application US/08991789A
; Patent No. 6225054
; GENERAL INFORMATION:
;     APPLICANT: Frudakis, Tony N.
;                Smith, John M.
;                Reed, Steven G.
;     TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;                         TREATMENT AND DIAGNOSIS OF BREAST CANCER
;     NUMBER OF SEQUENCES: 292
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE: Seed IP Law Group
;        STREET: 701 Fifth Avenue, Suite 6300
;        CITY: Seattle
;        STATE: Washington
;        COUNTRY: USA
;        ZIP: 98104-7092
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE: Floppy disk
;        COMPUTER: IBM PC compatible
;        OPERATING SYSTEM: PC-DOS/MS-DOS
;        SOFTWARE: PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/08/991,789A
;        FILING DATE: 11-Dec-1997
;        CLASSIFICATION: <Unknown>
;     ATTORNEY/AGENT INFORMATION:
;        NAME: Potter, Jane E. R.
;        REGISTRATION NUMBER: 33,332
;        REFERENCE/DOCKET NUMBER: 210121.419C3
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE: (206) 622-4900
;        TELEFAX: (206) 682-6031
;        ZIP: 98104-7092
; INFORMATION FOR SEQ ID NO: 222:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH: 578 base pairs
;        TYPE: nucleic acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
;     SEQUENCE DESCRIPTION: SEQ ID NO: 222:
US-08-991-789A-222

Alignment Scores:
Pred. No.:                   301        Length:          578
Score:                       8.00       Matches:           8
Percent Similarity:          100.0%     Conservative:      0
Best Local Similarity:       100.0%     Mismatches:        0
Query Match:                 3.8%       Indels:            0
DB:                          3          Gaps:              0

US-09-989-890-238 (1-212) x US-08-991-789A-222 (1-578)
```

```
Qy    32  SerArgProLeuSerProProPro 39
          |||||||||||||||||||||||
Db    153 TCGAGACCTCTGAGCCCACCTCCT 176

RESULT 66
US-09-062-451-222
; Sequence 222, Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
;     APPLICANT: Frudakis, Tony N.
;     APPLICANT: Smith, John M.
;     APPLICANT: Reed, Steven G.
;     TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;     TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
;     NUMBER OF SEQUENCES: 297
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE: SEED and BERRY LLP
;        STREET: 6300 Columbia Center, 701 Fifth Avenue
;        CITY: Seattle
;        STATE: Washington
;        COUNTRY: USA
;        ZIP: 98104-7092
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE: Floppy disk
;        COMPUTER: IBM PC compatible
;        OPERATING SYSTEM: PC-DOS/MS-DOS
;        SOFTWARE: PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/09/062,451
;        FILING DATE: 04-APR-1997
;        CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;        NAME: Maki, David J.
;        REGISTRATION NUMBER: 31,392
;        REFERENCE/DOCKET NUMBER: 210121.419C2
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE: (206) 622-4900
;        TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 222:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH: 578 base pairs
;        TYPE: nucleic acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
US-09-062-451-222

Alignment Scores:
Pred. No.:                   301        Length:          578
Score:                       8.00       Matches:           8
Percent Similarity:          100.0%     Conservative:      0
Best Local Similarity:       100.0%     Mismatches:        0
Query Match:                 3.8%       Indels:            0
DB:                          3          Gaps:              0

US-09-989-890-238 (1-212) x US-09-062-451-222 (1-578)

Qy    32  SerArgProLeuSerProProPro 39
          |||||||||||||||||||||||
Db    153 TCGAGACCTCTGAGCCCACCTCCT 176

RESULT 67
US-09-598-326-222
; Sequence 222, Application US/09598326
; Patent No. 6423496
; GENERAL INFORMATION:
;     APPLICANT: Frudakis, Tony N.
;                Smith, John M.
;                Reed, Steven G.
;     TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;                         TREATMENT AND DIAGNOSIS OF BREAST CANCER
;     NUMBER OF SEQUENCES: 247
;     CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Seed Intellectual Property Law Group PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/598,326
; FILING DATE: 20-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 222:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 222:
US-09-598-326-222
```

```
Alignment Scores:
Pred. No.:              301        Length:         578
Score:                  8.00       Matches:        8
Percent Similarity:     100.0%     Conservative:   0
Best Local Similarity:  100.0%     Mismatches:     0
Query Match:            3.8%       Indels:         0
DB:                     3          Gaps:           0

US-09-989-890-238 (1-212) x US-09-598-326-222 (1-578)

Qy   32 SerArgProLeuSerProPro 39
        ||||||||||||||||||||||
Db  153 TCGAGACCTCTGAGCCCACCTCCT 176
```

```
RESULT 68
US-09-289-198-222
; Sequence 222, Application US/09289198
; Patent No. 6586570
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; APPLICANT: Misher, Lynda
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C5
; CURRENT APPLICATION NUMBER: US/09/289,198
; CURRENT FILING DATE: 1999-04-09
; EARLIER APPLICATION NUMBER: US 09/062,451
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: US 08/991,789
; EARLIER FILING DATE: 1997-12-11
; EARLIER APPLICATION NUMBER: US 08/838,762
; EARLIER FILING DATE: 1997-04-09
; EARLIER APPLICATION NUMBER: PCT/US97/00485
; CURRENT FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: US 08/700,014
; EARLIER FILING DATE: 1996-08-20
; EARLIER APPLICATION NUMBER: US 08/585,392
; EARLIER FILING DATE: 1996-01-01
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSEQ for Windows Version 3.0
```

```
; SEQ ID NO 222
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(578)
; OTHER INFORMATION: n = A,T,C or G
US-09-289-198-222
```

```
Alignment Scores:
Pred. No.:              301        Length:         578
Score:                  8.00       Matches:        8
Percent Similarity:     100.0%     Conservative:   0
Best Local Similarity:  100.0%     Mismatches:     0
Query Match:            3.8%       Indels:         0
DB:                     3          Gaps:           0

US-09-989-890-238 (1-212) x US-09-289-198-222 (1-578)

Qy   32 SerArgProLeuSerProPro 39
        ||||||||||||||||||||||
Db  153 TCGAGACCTCTGAGCCCACCTCCT 176
```

```
RESULT 69
US-09-429-755-222
; Sequence 222, Application US/09429755A
; Patent No. 6656480
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; APPLICANT: Misher, Lynda
; APPLICANT: Retter, Marc W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C6
; CURRENT APPLICATION NUMBER: US/09/429,755A
; CURRENT FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 222
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(578)
; OTHER INFORMATION: n = A,T,C or G
US-09-429-755-222
```

```
Alignment Scores:
Pred. No.:              301        Length:         578
Score:                  8.00       Matches:        8
Percent Similarity:     100.0%     Conservative:   0
Best Local Similarity:  100.0%     Mismatches:     0
Query Match:            3.8%       Indels:         0
DB:                     3          Gaps:           0

US-09-989-890-238 (1-212) x US-09-429-755-222 (1-578)

Qy   32 SerArgProLeuSerProPro 39
        ||||||||||||||||||||||
Db  153 TCGAGACCTCTGAGCCCACCTCCT 176
```

```
RESULT 70
US-09-699-295-222
; Sequence 222, Application US/0969295
; Patent No. 6828431
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Reed, Steven G.
```

```
;       APPLICANT:  Smith, John M.
;       APPLICANT:  Misher, Linda E.
;       APPLICANT:  Dillon, Davin C.
;       APPLICANT:  Retter, Marc W.
;       APPLICANT:  Wang, Aijun
;       APPLICANT:  Skeiky, Yasir A.W.
;       APPLICANT:  Harlocker, Susan L.
;       TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
;       TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF BREAST CANCER
;       FILE REFERENCE:  210121.419C10
;       CURRENT APPLICATION NUMBER:  US/09/699,295
;       CURRENT FILING DATE:  2000-10-26
;       NUMBER OF SEQ ID NOS:  326
;       SOFTWARE:  FastSEQ for Windows Version 3.0
;       SEQ ID NO 222
;       LENGTH:  578
;       TYPE:  DNA
;       ORGANISM:  Homo sapien
;       FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  (1)...(578)
;       OTHER INFORMATION:  n = A,T,C or G
US-09-699-295-222

Alignment Scores:
Pred. No.:                301          Length:        578
Score:                   8.00          Matches:         8
Percent Similarity:     100.0%         Conservative:    0
Best Local Similarity:  100.0%         Mismatches:      0
Query Match:             3.8%          Indels:          0
DB:                         3          Gaps:            0

US-09-989-890-238 (1-212) x US-09-699-295-222 (1-578)

Qy    32 SerArgProLeuSerProProPro 39
            |||||||||||||||||||||||||
Db   153 TCGAGACCTCTGAGCCCACCTCCT 176

RESULT 71
US-09-534-825A-222
; Sequence 222, Application US/09534825A
; Patent No. 6861506
; GENERAL INFORMATION:
;       APPLICANT:  Frudakis, Tony N.
;       APPLICANT:  Reed, Steven G.
;       APPLICANT:  Smith, John M.
;       APPLICANT:  Misher, Lynda
;       APPLICANT:  Retter, Marc W.
;       APPLICANT:  Dillon, Davin C.
;       TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
;       TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF BREAST CANCER
;       FILE REFERENCE:  210121.419C7
;       CURRENT APPLICATION NUMBER:  US/09/534,825A
;       CURRENT FILING DATE:  2000-03-23
;       NUMBER OF SEQ ID NOS:  317
;       SOFTWARE:  FastSEQ for Windows Version 3.0
;       SEQ ID NO 222
;       LENGTH:  578
;       TYPE:  DNA
;       ORGANISM:  Homo sapien
;       FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  (1)...(578)
;       OTHER INFORMATION:  n = A,T,C or G
US-09-534-825A-222

Alignment Scores:
Pred. No.:                301          Length:        578
Score:                   8.00          Matches:         8
Percent Similarity:     100.0%         Conservative:    0
Best Local Similarity:  100.0%         Mismatches:      0
Query Match:             3.8%          Indels:          0
DB:                         3          Gaps:            0
```

```
US-09-989-890-238 (1-212) x US-09-534-825A-222 (1-578)

Qy    32 SerArgProLeuSerProPro 39
            |||||||||||||||||||||
Db   153 TCGAGACCTCTGAGCCCACCTCCT 176

RESULT 72
US-09-123-912-98
; Sequence 98, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
;       APPLICANT:  Reed, Steven G.
;       APPLICANT:  Wang, Tongtong
;       TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
;       FILE REFERENCE:  210121.455C1
;       CURRENT APPLICATION NUMBER:  US/09/123,912A
;       CURRENT FILING DATE:  1998-07-27
;       PRIOR APPLICATION NUMBER:  09/040,802
;       PRIOR FILING DATE:  1998-03-18
;       NUMBER OF SEQ ID NOS:  114
;       SOFTWARE:  PatentIn Ver. 2.0
;       SEQ ID NO 98
;       LENGTH:  600
;       TYPE:  DNA
;       ORGANISM:  Homo sapiens
;       FEATURE:
;       NAME/KEY:  modified_base
;       LOCATION:  (285)
;       OTHER INFORMATION:  Where n is a, c, g or t
;       NAME/KEY:  modified_base
;       LOCATION:  (349)
;       OTHER INFORMATION:  Where n is a, c, g or t
;       NAME/KEY:  modified_base
;       LOCATION:  (489)
;       OTHER INFORMATION:  Where n is a, c, g or t
;       NAME/KEY:  modified_base
;       LOCATION:  (496)
;       OTHER INFORMATION:  Where n is a, c, g or t
;       NAME/KEY:  modified_base
;       LOCATION:  (583)
;       OTHER INFORMATION:  Where n is a, c, g or t
US-09-123-912-98

Alignment Scores:
Pred. No.:                311          Length:        600
Score:                   8.00          Matches:         8
Percent Similarity:     100.0%         Conservative:    0
Best Local Similarity:  100.0%         Mismatches:      0
Query Match:             3.8%          Indels:          0
DB:                         3          Gaps:            0

US-09-989-890-238 (1-212) x US-09-123-912-98 (1-600)

Qy    55 ProSerAlaSerAlaAlaGly 62
            |||||||||||||||||||||
Db    79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 73
US-09-643-597-98
; Sequence 98, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
;       APPLICANT:  Wang, Tongtong
;       APPLICANT:  Fan, Liqun
;       APPLICANT:  Kalos, Michael D.
;       APPLICANT:  Bangur, Chaitanya S.
;       APPLICANT:  Hosken, Nancy
;       APPLICANT:  Fanger, Gary R.
;       APPLICANT:  Li, Samuel X.
;       APPLICANT:  Wang, Aijun
;       APPLICANT:  Skeiky, Yasir A.W.
;       APPLICANT:  Henderson, Robert A.
```

; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(600)
;   OTHER INFORMATION: n = A,T,C or G
US-09-643-597-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-643-597-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 74
US-09-480-884A-98
; Sequence 98, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C6
; CURRENT APPLICATION NUMBER: US/09/480,884A
; CURRENT FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 330
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(600)
;   OTHER INFORMATION: n = A,T,C or G
US-09-480-884A-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-480-884A-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 75
US-09-542-615A-98
; Sequence 98, Application US/09542615A
; Patent No. 6518256
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C8
; CURRENT APPLICATION NUMBER: US/09/542,615A
; CURRENT FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 350
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(600)
;   OTHER INFORMATION: n = A,T,C or G
US-09-542-615A-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-542-615A-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 76
US-09-606-421B-98
; Sequence 98, Application US/09606421B
; Patent No. 6531315
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C9
; CURRENT APPLICATION NUMBER: US/09/606,421B
; CURRENT FILING DATE: 2000-06-28
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(600)
;   OTHER INFORMATION: n = A,T,C or G
US-09-606-421B-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-606-421B-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 77
US-09-221-107-98
; Sequence 98, Application US/09221107
; Patent No. 6660838
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C2
; CURRENT APPLICATION NUMBER: US/09/221,107
; CURRENT FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 98
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (285)
; OTHER INFORMATION: Where n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (349)
; OTHER INFORMATION: Where n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (489)
; OTHER INFORMATION: Where n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (496)
; OTHER INFORMATION: Where n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (583)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-221-107-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-221-107-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 78
US-09-466-396A-98
; Sequence 98, Application US/09466396A
; Patent No. 6696247
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C4
; CURRENT APPLICATION NUMBER: US/09/466,396A
; CURRENT FILING DATE: 1999-12-17
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(600)
; OTHER INFORMATION: n = A,T,C or G
US-09-466-396A-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-466-396A-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 79
US-09-476-496A-98
; Sequence 98, Application US/09476496A
; Patent No. 6706262
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; TITLE OF INVENTION: LUNG CANCER
; FILE REFERENCE: 210121.455C5
; CURRENT APPLICATION NUMBER: US/09/476,496A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(600)
; OTHER INFORMATION: n = A,T,C or G
US-09-476-496A-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-476-496A-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 80
US-09-630-940B-98
; Sequence 98, Application US/09630940B

; Patent No. 6737514
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C10
; CURRENT APPLICATION NUMBER: US/09/630,940B
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(600)
; OTHER INFORMATION: n = A,T,C or G
US-09-630-940B-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-630-940B-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 81
US-09-285-479-98
; Sequence 98, Application US/09285479
; Patent No. 6821518
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF
; TITLE OF INVENTION: LUNG CANCER
; FILE REFERENCE: 210121.455C3
; CURRENT APPLICATION NUMBER: US/09/285,479
; CURRENT FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(600)
; OTHER INFORMATION: n = A,T,C or G
US-09-285-479-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-285-479-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 82
US-10-007-700-98
; Sequence 98, Application US/10007700
; Patent No. 6960570
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Li, Samuel X.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Fanger, Neil
; APPLICANT: Retter, Marc W.
; APPLICANT: Durham, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Peckman, David W.
; APPLICANT: Cai, Feng
; APPLICANT: Foy, Teresa M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C17
; CURRENT APPLICATION NUMBER: US/10/007,700
; CURRENT FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 295, 349, 489, 496, 583
; OTHER INFORMATION: n = A,T,C or G
US-10-007-700-98

Alignment Scores:
Pred. No.:   311   Length: 600
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-10-007-700-98 (1-600)

Qy   55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||
Db   79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 83
US-09-949-016-25231/c
; Sequence 25231, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25231
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-25231

Alignment Scores:
Pred. No.:              311          Length:        601
Score:                  8.00         Matches:       8
Percent Similarity:     100.0%       Conservative:  0
Best Local Similarity:  100.0%       Mismatches:    0
Query Match:            3.8%         Indels:        0
DB:                     3            Gaps:          0

US-09-989-890-238 (1-212) x US-09-949-016-25231 (1-601)

Qy    54 LeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||
Db   335 CTCCCCTCTGCAAGTGCCGCAGCT 312


RESULT 84
US-09-949-016-30352
; Sequence 30352, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30352
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-30352

Alignment Scores:
Pred. No.:              311          Length:        601
Score:                  8.00         Matches:       8
Percent Similarity:     100.0%       Conservative:  0
Best Local Similarity:  100.0%       Mismatches:    0
Query Match:            3.8%         Indels:        0
DB:                     3            Gaps:          0

US-09-989-890-238 (1-212) x US-09-949-016-30352 (1-601)

Qy    56 SerAlaSerAlaAlaAlaGlyIle 63
         ||||||||||||||||
Db   349 TCAGCCTCCGCAGCAGCTGGGATT 372


RESULT 85
US-09-949-016-33707
; Sequence 33707, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 33707
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-33707

Alignment Scores:
Pred. No.:              311          Length:        601
Score:                  8.00         Matches:       8
Percent Similarity:     100.0%       Conservative:  0
Best Local Similarity:  100.0%       Mismatches:    0
Query Match:            3.8%         Indels:        0
DB:                     3            Gaps:          0

US-09-989-890-238 (1-212) x US-09-949-016-33707 (1-601)

Qy    55 ProSerAlaSerAlaAlaGly 62
         |||||||||||||||||||||
Db   102 CCCTCAGCCTCAGCAGCAGCAGGT 125


RESULT 86
US-09-949-016-38879
; Sequence 38879, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38879
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-38879

Alignment Scores:
Pred. No.:              311          Length:        601
Score:                  8.00         Matches:       8
Percent Similarity:     100.0%       Conservative:  0
Best Local Similarity:  100.0%       Mismatches:    0
Query Match:            3.8%         Indels:        0
DB:                     3            Gaps:          0

US-09-989-890-238 (1-212) x US-09-949-016-38879 (1-601)

Qy    52 AlaGlyLeuProSerAlaSerAla 59
         ||||||||||||||||
Db   441 GCTGGGCTCCCGTCGGCGTCCGCC 464


RESULT 87
US-09-949-016-72530
```

```
Db        102 CCCTCAGCCTCAGCAGCAGCAGGT 125
              |||||||||||||||||||||||||
Db            CCCTCAGCCTCAGCAGCAGCAGGT 125

RESULT 89
US-09-949-016-88437
; Sequence 88437, Application US/0994901 6
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 88437
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-88437

Alignment Scores:
Pred. No.:                          Length:         601
Score:                   311        Matches:          8
Percent Similarity:      8.00       Conservative:     0
Best Local Similarity:   100.0%     Mismatches:       0
Query Match:             3.8%       Indels:           0
DB:                      3          Gaps:             0

US-09-989-890-238 (1-212) x US-09-949-016-88437 (1-601)

Qy         52 AlaGlyLeuProSerAlaSerAla 59
              ||||||||||||||||||||||||
Db        441 GCTGGGCTCCCGTCGGCGTCCGCC 464

RESULT 90
US-09-949-016-92611
; Sequence 92611, Application US/0994901 6
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 92611
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-92611

Alignment Scores:
Pred. No.:                          Length:         601
Score:                   311        Matches:          8
Percent Similarity:      8.00       Conservative:     0
Best Local Similarity:   100.0%     Mismatches:       0
Query Match:             3.8%       Indels:           0
```

```
; Sequence 72530, Application US/0994901 6
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72530
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-72530

Alignment Scores:
Pred. No.:                          Length:         601
Score:                   311        Matches:          8
Percent Similarity:      8.00       Conservative:     0
Best Local Similarity:   100.0%     Mismatches:       0
Query Match:             3.8%       Indels:           0
DB:                      3          Gaps:             0

US-09-989-890-238 (1-212) x US-09-949-016-72530 (1-601)

Qy         32 SerArgProLeuSerProProPro 39
              ||||||||||||||||||||||||
Db        557 TCCCGGCCCCTCTCCCCTCCCCCA 580

RESULT 88
US-09-949-016-76311
; Sequence 76311, Application US/0994901 6
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 76311
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-76311

Alignment Scores:
Pred. No.:                          Length:         601
Score:                   311        Matches:          8
Percent Similarity:      8.00       Conservative:     0
Best Local Similarity:   100.0%     Mismatches:       0
Query Match:             3.8%       Indels:           0
DB:                      3          Gaps:             0

US-09-989-890-238 (1-212) x US-09-949-016-76311 (1-601)

Qy         55 ProSerAlaSerAlaAlaAlaGly 62
```

DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-949-016-92611 (1-601)

Qy   27 SerLeuAlaLeuLeuSerArgPro 34
             ||||||||||||||||||||||||
Db   137 AGTCTTGCTCTGTTGTCCAGGCCG 160

RESULT 91
US-09-949-016-92612
; Sequence 92612, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-09-08
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-10-03
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 92612
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-92612

Alignment Scores:
Pred. No.:   311   Length: 601
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-949-016-92612 (1-601)

Qy   27 SerLeuAlaLeuLeuSerArgPro 34
             ||||||||||||||||||||||||
Db   180 AGTCTTGCTCTGTTGTCCAGGCCG 203

RESULT 92
US-09-949-016-92613
; Sequence 92613, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-09-08
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-10-03
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 92613
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-92613

Alignment Scores:

Pred. No.:   311   Length: 601
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-949-016-92613 (1-601)

Qy   27 SerLeuAlaLeuLeuSerArgPro 34
             ||||||||||||||||||||||||
Db   320 AGTCTTGCTCTGTTGTCCAGGCCG 343

RESULT 93
US-09-949-016-123793
; Sequence 123793, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-09-08
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 123793
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-123793

Alignment Scores:
Pred. No.:   311   Length: 601
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-949-016-123793 (1-601)

Qy   56 SerAlaSerAlaAlaAlaGlyIle 63
             ||||||||||||||||||||||||
Db   349 TCAGCCTCCGCAGCAGCTGGGATT 372

RESULT 94
US-09-949-016-133788/c
; Sequence 133788, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 133788
;   LENGTH: 601

```
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-133788

Alignment Scores:
Pred. No.:                        311        Length:              601
Score:                            8.00       Matches:             8
Percent Similarity:               100.0%     Conservative:        0
Best Local Similarity:            100.0%     Mismatches:          0
Query Match:                      3.8%       Indels:              0
DB:                               3          Gaps:                0

US-09-989-890-238 (1-212) x US-09-949-016-133788 (1-601)

Qy    24 LeuProSerSerLeuAlaLeuLeu 31
         ||||||||||||||||||||||||
Db   534 CTACCTTCCAGTCTTGCTCTTCTT 511

RESULT 95
US-09-949-016-133789/c
;   Sequence 133789, Application US/09949016
;   Patent No. 6812339
;   GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 133789
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-133789

Alignment Scores:
Pred. No.:                        311        Length:              601
Score:                            8.00       Matches:             8
Percent Similarity:               100.0%     Conservative:        0
Best Local Similarity:            100.0%     Mismatches:          0
Query Match:                      3.8%       Indels:              0
DB:                               3          Gaps:                0

US-09-989-890-238 (1-212) x US-09-949-016-133789 (1-601)

Qy    24 LeuProSerSerLeuAlaLeuLeu 31
         ||||||||||||||||||||||||
Db   516 CTACCTTCCAGTCTTGCTCTTCTT 493

RESULT 96
US-09-949-016-141154
;   Sequence 141154, Application US/09949016
;   Patent No. 6812339
;   GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 141154
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-141154

Alignment Scores:
Pred. No.:                        311        Length:              601
Score:                            8.00       Matches:             8
Percent Similarity:               100.0%     Conservative:        0
Best Local Similarity:            100.0%     Mismatches:          0
Query Match:                      3.8%       Indels:              0
DB:                               3          Gaps:                0

US-09-989-890-238 (1-212) x US-09-949-016-141154 (1-601)

Qy    55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||||
Db   484 CCTTCAGCCTCAGCGGCTGCCGGA 507

RESULT 97
US-09-949-016-141155
;   Sequence 141155, Application US/09949016
;   Patent No. 6812339
;   GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 141155
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-141155

Alignment Scores:
Pred. No.:                        311        Length:              601
Score:                            8.00       Matches:             8
Percent Similarity:               100.0%     Conservative:        0
Best Local Similarity:            100.0%     Mismatches:          0
Query Match:                      3.8%       Indels:              0
DB:                               3          Gaps:                0

US-09-989-890-238 (1-212) x US-09-949-016-141155 (1-601)

Qy    55 ProSerAlaSerAlaAlaAlaGly 62
         ||||||||||||||||||||||||
Db   202 CCTTCAGCCTCAGCGGCTGCCGGA 225

RESULT 98
US-09-949-016-147427/c
;   Sequence 147427, Application US/09949016
;   Patent No. 6812339
;   GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
```

```
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; SOFTWARE: FastSEQ for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 207012
; SEQ ID NO 147427
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-147427

Alignment Scores:
Pred. No.:               311              Length:         601
Score:                   8.00             Matches:        8
Percent Similarity:      100.0%           Conservative:   0
Best Local Similarity:   100.0%           Mismatches:     0
Query Match:             3.8%             Indels:         0
DB:                      3                Gaps:           0

US-09-989-890-238 (1-212) x US-09-949-016-147427 (1-601)

Qy    54 LeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||
Db   335 CTCCCCTCTGCAAGTGCCGCAGCT 312
```

RESULT 99
US-09-949-016-147591
```
; Sequence 147591, Application US/0994016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   NUMBER OF SEQ ID NOS: 207012
; SEQ ID NO 147591
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-147591

Alignment Scores:
Pred. No.:               311              Length:         601
Score:                   8.00             Matches:        8
Percent Similarity:      100.0%           Conservative:   0
Best Local Similarity:   100.0%           Mismatches:     0
Query Match:             3.8%             Indels:         0
DB:                      3                Gaps:           0

US-09-989-890-238 (1-212) x US-09-949-016-147591 (1-601)

Qy    56 SerAlaSerAlaAlaAlaGlyIle 63
         ||||||||||||||||||||||||
Db   349 TCAGCCTCCGCAGCAGCTGGGATT 372
```

RESULT 100
US-09-949-016-157906
```
; Sequence 157906, Application US/0994016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   NUMBER OF SEQ ID NOS: 207012
; SEQ ID NO 157906
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-157906

Alignment Scores:
Pred. No.:               311              Length:         601
Score:                   8.00             Matches:        8
Percent Similarity:      100.0%           Conservative:   0
Best Local Similarity:   100.0%           Mismatches:     0
Query Match:             3.8%             Indels:         0
DB:                      3                Gaps:           0

US-09-989-890-238 (1-212) x US-09-949-016-157906 (1-601)

Qy    25 ProSerSerLeuAlaLeuLeuSer 32
         ||||||||||||||||||||||||
Db   477 CCATCCAGCTTGGCTCTGCTGTCG 500

Search completed: March 17, 2006, 06:38:21
Job time : 822 secs
```

OM protein - nucleic search, using frame_plus_p2n model

```
              Copyright (c) 1993 - 2006  Biocceleration Ltd.
```

GenCore version 5.1.7

Run on: March 17, 2006, 06:27:16 ; Search time 775 Seconds
                                          (without alignments)
                                          2262.075 Million cell updates/sec

Title: US-09-989-890-238
Perfect score: 212
Sequence: 212 1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table: OLIGO

```
     Xgapop   60.0 , Xgapext  60.0
     Ygapop   60.0 , Ygapext  60.0
     Fgapop    6.0 , Fgapext   7.0
     Delop     6.0 , Delext    7.0
```

Searched: 9793542 seqs, 4134689005 residues

Word size: 1

Total number of hits satisfying chosen parameters: 19571230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095520_17643/app_query.fasta_1
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=oligo_p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=500 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=100 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -HOST=abss03h
-USER=US09989890 @CGN_1_1026 @runat_16032006_095520_17643 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :   Published Applications NA Main:*
```
     1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
     2:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
     3:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
     4:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
     5:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
     6:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
     7:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
     8:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
     9:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
    10:  /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	212	100.0	1713	3	US-09-989-890-105	Sequence 105, App
2	115	54.2	1785	6	US-10-104-047-799	Sequence 799, App
3	81	38.2	427	9	US-10-779-543-12051	Sequence 12051, A
4	69	32.5	654	5	US-10-074-475-124	Sequence 124, Appl
5	59	27.8	1977	3	US-09-989-920-56	Sequence 56, Appl
6	59	27.8	1977	3	US-09-989-920-117	Sequence 117, App
7	11	5.2	253861	7	US-10-741-601-5611	Sequence 5611, Ap

	8		11	5.2	261817	5	US-10-087-192-2002	Sequence 2002, Ap
	9		10	4.7	523	5	US-10-027-632-143483	Sequence 143483,
	10		10	4.7	523	5	US-10-027-632-143484	Sequence 143484,
	11		10	4.7	523	6	US-10-027-632-143483	Sequence 143483,
	12		10	4.7	523	6	US-10-027-632-143484	Sequence 143484,
	13	c	10	4.7	861	6	US-10-156-761-4893	Sequence 4893, Ap
	14		10	4.7	5723	6	US-10-007-926A-134	Sequence 134, App
	15		10	4.7	5723	6	US-10-341-434-52	Sequence 52, Appl
	16		10	4.7	5723	7	US-10-671-074-4	Sequence 4, Appli
	17		10	4.7	5723	7	US-10-956-157-760	Sequence 760, App
	18	c	10	4.7	5833	9	US-10-450-763-20857	Sequence 20857, A
	19	c	10	4.7	9025608	6	US-10-456-761-1	Sequence 1, Appli
	20	c	9	4.2	201	8	US-10-719-993-6100	Sequence 6100, Ap
	21	c	9	4.2	201	8	US-10-719-993-6190	Sequence 6190, Ap
	22		9	4.2	201	8	US-10-719-993-6280	Sequence 6280, Ap
	23	c	9	4.2	201	8	US-10-719-993-6416	Sequence 6416, Ap
	24	c	9	4.2	201	8	US-10-719-993-39711	Sequence 39711, A
	25	c	9	4.2	239	6	US-10-243-552-741	Sequence 741, App
	26	c	9	4.2	239	6	US-10-450-763-10524	Sequence 10524, A
	27		9	4.2	496	3	US-09-918-995-3937	Sequence 3937, Ap
	28	c	9	4.2	546	6	US-10-195-730-73	Sequence 73, Appl
	29	c	9	4.2	546	7	US-10-799-747-73	Sequence 73, Appl
	30	c	9	4.2	546	7	US-10-979-183-73	Sequence 73, Appl
	31	c	9	4.2	549	3	US-09-764-875-478	Sequence 478, App
	32	c	9	4.2	562	8	US-10-264-237-488	Sequence 488, App
	33	c	9	4.2	568	7	US-10-437-963-84925	Sequence 84925, A
	34		9	4.2	637	4	US-09-925-065A-468689	Sequence 468689,
	35		9	4.2	645	7	US-10-437-963-14833	Sequence 14833, A
	36		9	4.2	652	4	US-09-925-065A-726141	Sequence 726141,
	37		9	4.2	717	7	US-10-260-238-4042	Sequence 4042, Ap
	38	c	9	4.2	737	7	US-10-260-238-4040	Sequence 4040, Ap
	39		9	4.2	792	7	US-10-296-115-280	Sequence 280, App
	40	c	9	4.2	824	7	US-10-260-238-4041	Sequence 4041, Ap
	41		9	4.2	891	7	US-10-437-963-86193	Sequence 86193, A
	42	c	9	4.2	963	7	US-10-437-963-34733	Sequence 34733, A
	43		9	4.2	1209	6	US-10-087-080-31	Sequence 31, Appl
	44	c	9	4.2	1209	7	US-10-229-345-9	Sequence 9, Appli
	45		9	4.2	1209	7	US-10-274-177-9	Sequence 9, Appli
	46		9	4.2	1209	7	US-10-650-112-9	Sequence 9, Appli
	47	c	9	4.2	1212	10	US-11-041-788-9	Sequence 9, Appli
	48	c	9	4.2	1212	7	US-10-650-112-25	Sequence 25, Appl
	49	c	9	4.2	1448	9	US-10-756-149-15	Sequence 15, Appl
	50		9	4.2	1564	3	US-09-764-875-179	Sequence 179, App
	51	c	9	4.2	1644	6	US-10-471-010-1	Sequence 61, Appl
	52		9	4.2	1818	7	US-10-332-447-61	Sequence 90203, A
	53		9	4.2	1923	7	US-10-437-963-90203	Sequence 26078, A
	54	c	9	4.2	2071	6	US-10-369-493-26078	Sequence 62, Appl
	55		9	4.2	2245	7	US-10-332-447-62	Sequence 15603, A
	56	c	9	4.2	2256	7	US-10-450-763-15603	Sequence 4805, Ap
	57		9	4.2	2403	7	US-10-437-963-4805	Sequence 691773,
	58		9	4.2	2520	4	US-09-925-065A-691773	Sequence 86961, A
	59		9	4.2	3795	7	US-10-437-963-86961	Sequence 23948, A
	60		9	4.2	3840	10	US-11-097-143-23948	Sequence 23947, A
	61	c	9	4.2	5906	10	US-11-097-143-23947	Sequence 960, App
	62	c	9	4.2	7880	3	US-09-764-860-960	Sequence 961, App
	63	c	9	4.2	7880	3	US-09-764-860-961	Sequence 960, App
	64	c	9	4.2	7880	5	US-10-074-095-960	Sequence 960, App
	65	c	9	4.2	7880	6	US-10-212-872-960	Sequence 961, App
	66	c	9	4.2	7880	6	US-10-212-872-961	Sequence 961, App
	67		9	4.2	7880	6	US-10-156-761-411	Sequence 411, App
	68		9	4.2	10056	6	US-10-919-172-97	Sequence 97, Appl
	69		9	4.2	10432	3	US-09-919-039-20	Sequence 20, Appl
	70		9	4.2	10432	5	US-10-084-817-349	Sequence 349, App
	71		9	4.2	10432	8	US-10-133-013-262	Sequence 262, App
	72		9	4.2	10432	8	US-10-752-996-97	Sequence 97, Appl
	73		9	4.2	11466	8	US-10-723-860-5608	Sequence 5608, Ap
	74	c	9	4.2	11598	8	US-10-719-993-402	Sequence 402, App
	75	c	9	4.2	11602	8	US-10-719-993-403	Sequence 403, App
	76		9	4.2	12493	8	US-10-739-946-1	Sequence 1, Appli
	77	c	9	4.2	12494	5	US-10-252-157-263	Sequence 263, App
	78		9	4.2	12505	6	US-10-115-081-18	Sequence 18, Appl
	79	c	9	4.2	12515	3	US-09-918-624B-48	Sequence 48, Appl

Flag	No.			Number	Patent	n	Sequence
c	81	9	4.2	12515	US-09-968-007A-989	3	Sequence 989, App
c	82	9	4.2	12515	US-10-388-360-313	6	Sequence 313, App
	83	9	4.2	12515	US-10-719-993-405	8	Sequence 405, App
	84	9	4.2	12515	US-10-843-641A-7459	9	Sequence 7459, App
	85	9	4.2	12515	US-10-928-465-58	5	Sequence 58, Appl
	86	9	4.2	12632	US-10-071-766-49	9	Sequence 49, Appl
	87	9	4.2	12678	US-10-719-993-404	8	Sequence 404, App
	88	9	4.2	41736	US-10-719-993-6886	8	Sequence 6886, Ap
	89	9	4.2	48396	US-10-162-846-11	6	Sequence 11, Appl
	90	9	4.2	73544	US-10-471-010-3	8	Sequence 3, Appli
c	91	9	4.2	96600	US-09-997-722-85	3	Sequence 85, Appl
	92	9	4.2	100000	US-10-156-761-15103	6	Sequence 15103, A
	93	9	4.2	183046	US-10-981-277-31	8	Sequence 31, Appl
	94	9	4.2	493999	US-10-719-993-6787	8	Sequence 6787, Ap
	95	9	4.2	9025608	US-10-156-761-1	6	Sequence 1, Appli
c	96	8	3.8	25	US-09-719-900-172651	10	Sequence 172651,
c	97	8	3.8	25	US-11-036-317-669570	10	Sequence 669570,
	98	8	3.8	25	US-11-036-317-669571	10	Sequence 669571,
	99	8	3.8	25	US-11-036-317-687056	6	Sequence 687056,
	100	8	3.8	50	US-10-131-827-69	6	Sequence 69, Appl
	101	8	3.8	50	US-10-131-827-4540	6	Sequence 4540, Ap
	102	8	3.8	55	US-09-989-339-42	3	Sequence 42, Appl
c	103	8	3.8	55	US-09-989-339-43	5	Sequence 43, Appl
	104	8	3.8	55	US-10-023-066A-96	5	Sequence 96, Appl
c	105	8	3.8	55	US-10-023-066A-97	8	Sequence 97, Appl
	106	8	3.8	55	US-10-804-678-96	9	Sequence 96, Appl
	107	8	3.8	55	US-10-804-678-97	8	Sequence 97, Appl
	108	8	3.8	144	US-10-184-085A-1283	7	Sequence 1283, Ap
	109	8	3.8	167	US-10-424-599-107733	9	Sequence 107733,
	110	8	3.8	179	US-10-029-386-23120	6	Sequence 23120, A
	111	8	3.8	186	US-09-864-761-22513	9	Sequence 22513, A
	112	8	3.8	192	US-10-921-023-16	8	Sequence 16, Appl
	113	8	3.8	201	US-10-252-157-337	6	Sequence 337, App
	114	8	3.8	201	US-10-741-600-18451	6	Sequence 18451, A
	115	8	3.8	220	US-10-029-386-27137	6	Sequence 27137, A
c	116	8	3.8	227	US-09-728-446-1439	5	Sequence 1439, Ap
	117	8	3.8	274	US-10-040-739-175	8	Sequence 175, App
	118	8	3.8	282	US-10-437-963-39690	7	Sequence 39690, A
c	119	8	3.8	314	US-10-437-963-68026	7	Sequence 68026, A
	120	8	3.8	317	US-09-925-065A-620441	6	Sequence 620441,
	121	8	3.8	340	US-10-429-802-28	9	Sequence 28, Appl
	122	8	3.8	340	US-10-430-503-19	6	Sequence 19, Appl
	123	8	3.8	340	US-10-928-626-3	6	Sequence 3, Appli
	124	8	3.8	340	US-11-010-980A-3	10	Sequence 3, Appli
	125	8	3.8	349	US-09-969-708-179	9	Sequence 179, App
	126	8	3.8	349	US-10-843-641A-7650	9	Sequence 7650, Ap
	127	8	3.8	357	US-10-437-963-7356	7	Sequence 7356, Ap
	128	8	3.8	358	US-10-425-115-179373	8	Sequence 179373,
	129	8	3.8	363	US-10-767-701-23397	9	Sequence 23397, A
	130	8	3.8	369	US-10-915-740A-601	9	Sequence 601, App
c	131	8	3.8	377	US-09-925-065A-363884	4	Sequence 363884,
	132	8	3.8	377	US-09-925-065A-363885	5	Sequence 363885,
	133	8	3.8	377	US-09-925-065A-363886	4	Sequence 363886,
	134	8	3.8	382	US-09-954-456-1566	3	Sequence 1566, Ap
	135	8	3.8	382	US-09-880-107-2973	3	Sequence 2973, Ap
	136	8	3.8	382	US-09-954-531-120	9	Sequence 120, App
c	137	8	3.8	382	US-09-954-531-344	9	Sequence 344, App
	138	8	3.8	382	US-10-843-641A-1187	9	Sequence 1187, Ap
	139	8	3.8	382	US-10-843-641A-1411	7	Sequence 1411, Ap
c	140	8	3.8	382	US-10-843-641A-4593	3	Sequence 4593, Ap
	141	8	3.8	384	US-10-437-963-14121	7	Sequence 14121, A
	142	8	3.8	392	US-10-425-115-36301	8	Sequence 36301, A
c	143	8	3.8	392	US-10-425-115-71200	9	Sequence 71200, A
	144	8	3.8	393	US-09-918-995-35329	7	Sequence 35329, A
	145	8	3.8	396	US-10-437-963-84697	9	Sequence 84697, A
	146	8	3.8	397	US-09-925-065A-13536	4	Sequence 13536, A
	147	8	3.8	408	US-09-925-065A-844771	4	Sequence 844771,
	148	8	3.8	411	US-09-814-353-14673	3	Sequence 14673, A
c	149	8	3.8	428	US-09-918-995-23488	3	Sequence 23488, A
	150	8	3.8	429	US-09-918-995-35774	3	Sequence 35774, A
c	151	8	3.8	435	US-10-282-122A-6867	7	Sequence 6867, Ap
	152	8	3.8	436	US-10-357-930-39081	8	Sequence 39081, App
c	153	8	3.8	441	US-09-735-705-97	3	Sequence 97, Appl

Flag	No.			Number	n	Patent	Sequence
	154	8	3.8	441	3	US-09-850-716A-97	Sequence 97, Appl
	155	8	3.8	441	3	US-09-897-778-97	Sequence 97, Appl
	156	8	3.8	441	3	US-09-796-692-3755	Sequence 3755, Ap
	157	8	3.8	441	3	US-09-466-396A-97	Sequence 97, Appl
	158	8	3.8	441	5	US-10-007-700-97	Sequence 97, Appl
	159	8	3.8	441	6	US-10-040-862-3755	Sequence 3755, Ap
	160	8	3.8	441	6	US-10-117-982-97	Sequence 97, Appl
	161	8	3.8	441	6	US-10-313-986-97	Sequence 97, Appl
	162	8	3.8	441	6	US-10-057-475B-3755	Sequence 3755, Ap
	163	8	3.8	441	6	US-10-154-884B-3755	Sequence 3755, Ap
	164	8	3.8	441	8	US-10-764-324-3755	Sequence 3755, Ap
	165	8	3.8	441	9	US-10-775-972-97	Sequence 97, Appl
	166	8	3.8	441	3	US-10-922-124-97	Sequence 97, Appl
c	167	8	3.8	442	3	US-09-864-761-5751	Sequence 5751, Ap
	168	8	3.8	445	3	US-10-637-855-384	Sequence 384, App
	169	8	3.8	451	3	US-09-918-995-2661	Sequence 2661, Ap
	170	8	3.8	464	8	US-10-425-115-90782	Sequence 90782, A
c	171	8	3.8	464	8	US-10-357-930-14970	Sequence 14970, A
c	172	8	3.8	467	3	US-09-918-995-26473	Sequence 26473, A
c	173	8	3.8	468	7	US-10-282-122A-37355	Sequence 37355, A
	174	8	3.8	473	3	US-09-734-569-51	Sequence 51, Appl
c	175	8	3.8	473	4	US-09-925-065A-489934	Sequence 489934,
	176	8	3.8	474	3	US-09-918-995-29483	Sequence 29483, A
	177	8	3.8	488	8	US-10-278-698-496	Sequence 496, App
c	178	8	3.8	488	8	US-10-278-698-1012	Sequence 1012, Ap
	179	8	3.8	489	7	US-10-437-963-43545	Sequence 43545, A
c	180	8	3.8	492	8	US-10-278-698-492	Sequence 492, App
c	181	8	3.8	492	8	US-10-278-698-1008	Sequence 1008, Ap
c	182	8	3.8	493	8	US-09-918-995-15385	Sequence 15385, A
	183	8	3.8	493	8	US-10-357-930-16189	Sequence 16189, A
	184	8	3.8	500	7	US-10-437-963-102240	Sequence 102240,
	185	8	3.8	500	8	US-10-425-115-101696	Sequence 101696,
c	186	8	3.8	501	8	US-10-425-115-45779	Sequence 45779, A
	187	8	3.8	503	7	US-10-437-963-14186	Sequence 14186, A
	188	8	3.8	507	3	US-10-156-761-3733	Sequence 3733, Ap
c	189	8	3.8	508	7	US-10-972-079-88114	Sequence 88114, A
c	190	8	3.8	508	3	US-09-815-343-1278	Sequence 1278, Ap
c	191	8	3.8	509	7	US-10-097-105-1278	Sequence 1278, Ap
c	192	8	3.8	509	3	US-10-012-697-701	Sequence 701, App
	193	8	3.8	509	7	US-10-767-701-3215	Sequence 3215, App
	194	8	3.8	512	6	US-10-779-543-22701	Sequence 22701, A
	195	8	3.8	512	9	US-10-972-079-88113	Sequence 88113, A
	196	8	3.8	513	9	US-10-029-386-9420	Sequence 9420, Ap
	197	8	3.8	514	7	US-10-437-963-96325	Sequence 96325, A
c	198	8	3.8	516	4	US-09-925-065A-616675	Sequence 616675,
	199	8	3.8	517	5	US-09-925-065A-616676	Sequence 616676,
	200	8	3.8	517	9	US-10-972-079-9026	Sequence 9026, Ap
c	201	8	3.8	518	9	US-09-925-065A-418469	Sequence 418469,
c	202	8	3.8	522	6	US-10-320-800-3	Sequence 1, Appli
	203	8	3.8	525	7	US-10-650-123-1	Sequence 1, Appli
	204	8	3.8	529	9	US-09-796-692-9257	Sequence 9257, Ap
	205	8	3.8	529	5	US-10-040-862-9257	Sequence 9257, Ap
	206	8	3.8	529	6	US-10-057-475B-9257	Sequence 9257, Ap
	207	8	3.8	529	8	US-10-154-884B-9257	Sequence 9257, Ap
	208	8	3.8	531	6	US-10-764-324-9257	Sequence 128, App
	209	8	3.8	532	8	US-10-243-552-128	Sequence 1471, Ap
c	210	8	3.8	532	8	US-10-363-345A-1471	Sequence 1472, Ap
	211	8	3.8	532	9	US-10-363-345A-1472	Sequence 1471, Ap
	212	8	3.8	545	6	US-10-363-483A-1471	Sequence 1472, Ap
	213	8	3.8	545	7	US-10-363-483A-1472	Sequence 22849, A
	214	8	3.8	547	9	US-10-029-386-22849	Sequence 204, App
	215	8	3.8	547	4	US-10-621-901-204	Sequence 4173, Ap
	216	8	3.8	547	9	US-09-796-692-4173	Sequence 151898,
c	217	8	3.8	547	5	US-09-925-065A-151898	Sequence 4173, Ap
	218	8	3.8	547	6	US-10-040-862-4173	Sequence 4173, Ap
	219	8	3.8	547	8	US-10-057-475B-4173	Sequence 4173, Ap
	220	8	3.8	548	8	US-10-154-884B-4173	Sequence 4173, Ap
	221	8	3.8	548	9	US-10-764-324-4173	Sequence 162350,
	222	8	3.8	549	6	US-10-425-115-162350	Sequence 281060,
c	223	8	3.8	549	9	US-10-027-632-281060	Sequence 12524, A
c	224	8	3.8	549	6	US-10-029-386-12524	Sequence 281060,
	225	8	3.8	549	9	US-10-027-632-281060	Sequence 3, Appli
c	226	8	3.8	549	7	US-10-425-037-3	Sequence 3, Appli

						US identifier	Sequence
c	227	3.8	8	552	4	US-09-925-065A-748081	Sequence 748081,
c	228	3.8	8	559	4	US-09-925-065A-808255	Sequence 808255,
c	229	3.8	8	559	4	US-09-925-065A-808256	Sequence 808256,
c	230	3.8	8	559	4	US-09-925-065A-808257	Sequence 808257,
c	231	3.8	8	562	4	US-10-027-632-192118	Sequence 192118,
	232	3.8	8	562	6	US-10-027-632-192118	Sequence 192118,
	233	3.8	8	563	6	US-10-029-386-3613	Sequence 3613, Ap
	234	3.8	8	565	5	US-09-925-065A-946604	Sequence 946604,
	235	3.8	8	565	5	US-10-027-632-270152	Sequence 270152,
c	236	3.8	8	565	6	US-10-027-632-270153	Sequence 270153,
c	237	3.8	8	565	5	US-10-027-632-270152	Sequence 270152,
	238	3.8	8	565	5	US-10-027-632-270153	Sequence 270153,
c	239	3.8	8	567	6	US-10-027-632-187680	Sequence 187680,
c	240	3.8	8	567	5	US-10-027-632-187680	Sequence 187680,
	241	3.8	8	568	7	US-10-021-323-11313	Sequence 11313, A
	242	3.8	8	570	4	US-09-925-065A-693639	Sequence 693639,
c	243	3.8	8	570	4	US-09-925-065A-950741	Sequence 950741,
	244	3.8	8	571	4	US-09-925-065A-733926	Sequence 733926,
c	245	3.8	8	571	6	US-10-029-386-2244	Sequence 2244, Ap
	246	3.8	8	576	6	US-09-925-065A-854936	Sequence 854936,
	247	3.8	8	576	4	US-10-029-386-11942	Sequence 11942, A
	248	3.8	8	577	3	US-09-864-761-6755	Sequence 6755, Ap
	249	3.8	8	578	3	US-09-810-936-222	Sequence 222, App
	250	3.8	8	578	4	US-09-429-755-222	Sequence 222, App
	251	3.8	8	578	4	US-09-924-400-222	Sequence 222, App
	252	3.8	8	578	4	US-10-212-679-222	Sequence 222, App
	253	3.8	8	578	7	US-10-079-137B-222	Sequence 222, App
	254	3.8	8	578	7	US-10-425-115-70375	Sequence 70375, A
	255	3.8	8	579	8	US-10-198-846-9350	Sequence 9350, Ap
	256	3.8	8	579	7	US-09-796-692-115704	Sequence 115704,
	257	3.8	8	581	4	US-09-925-065A-53760	Sequence 53760, A
c	258	3.8	8	581	4	US-09-925-065A-53761	Sequence 53761, A
c	259	3.8	8	581	4	US-09-925-065A-53762	Sequence 53762, A
c	260	3.8	8	582	7	US-10-767-701-674	Sequence 674, App
c	261	3.8	8	585	5	US-10-027-632-170836	Sequence 170836,
c	262	3.8	8	585	5	US-10-027-632-170836	Sequence 170836,
c	263	3.8	8	585	7	US-10-437-963-62813	Sequence 62813, A
c	264	3.8	8	586	5	US-09-925-065A-295614	Sequence 295614,
c	265	3.8	8	589	6	US-10-029-386-13437	Sequence 13437, A
c	266	3.8	8	591	3	US-09-796-692-5847	Sequence 5847, Ap
	267	3.8	8	591	3	US-10-942-583-12	Sequence 12, Appl
c	268	3.8	8	591	5	US-10-040-862-5847	Sequence 5847, Ap
c	269	3.8	8	591	6	US-10-057-475B-5847	Sequence 5847, Ap
c	270	3.8	8	591	7	US-10-154-884B-5847	Sequence 5847, Ap
	271	3.8	8	591	6	US-10-437-963-26086	Sequence 26086, A
	272	3.8	8	591	8	US-10-764-324-5847	Sequence 5847, Ap
	273	3.8	8	593	4	US-10-425-115-129854	Sequence 129854,
c	274	3.8	8	593	4	US-09-925-065A-630739	Sequence 630739,
c	275	3.8	8	595	4	US-09-925-065A-295459	Sequence 295459,
c	276	3.8	8	595	4	US-09-925-065A-641425	Sequence 641425,
c	277	3.8	8	599	4	US-09-925-065A-301227	Sequence 301227,
c	278	3.8	8	599	4	US-09-925-065A-301228	Sequence 301228,
	279	3.8	8	599	3	US-10-972-079-88109	Sequence 88109, A
	280	3.8	8	599	3	US-10-972-079-90063	Sequence 90063, A
	281	3.8	8	599	9	US-10-972-079-92901	Sequence 92901, A
	282	3.8	8	600	9	US-09-735-705-98	Sequence 98, Appl
	283	3.8	8	600	3	US-09-850-716A-98	Sequence 98, Appl
	284	3.8	8	600	3	US-09-897-778-98	Sequence 98, Appl
	285	3.8	8	600	5	US-09-466-396A-98	Sequence 98, Appl
	286	3.8	8	600	6	US-10-007-700-98	Sequence 98, Appl
	287	3.8	8	600	6	US-10-117-982-98	Sequence 98, Appl
	288	3.8	8	600	6	US-10-313-986-98	Sequence 98, Appl
	289	3.8	8	600	9	US-10-775-972-98	Sequence 98, Appl
	290	3.8	8	600	9	US-10-922-124-98	Sequence 98, Appl
	291	3.8	8	600	9	US-10-972-079-8538	Sequence 8538, Ap
	292	3.8	8	600	9	US-10-972-079-9024	Sequence 9024, Ap
c	293	3.8	8	600	9	US-10-972-079-9025	Sequence 9025, Ap
c	294	3.8	8	600	9	US-10-972-079-88002	Sequence 88002, A
c	295	3.8	8	600	9	US-10-972-079-88110	Sequence 88110, A
c	296	3.8	8	600	9	US-10-972-079-88111	Sequence 88111, A
c	297	3.8	8	600	9	US-10-972-079-88112	Sequence 88112, A
c	298	3.8	8	600	9	US-10-972-079-92902	Sequence 92902, A
c	299	3.8	8	600	9	US-10-972-079-92903	Sequence 92903, A

						US identifier	Sequence
c	300	8	3.8	602	4	US-09-925-065A-774077	Sequence 774077,
c	301	8	3.8	603	4	US-09-925-065A-837176	Sequence 837176,
c	302	8	3.8	603	4	US-09-925-065A-571840	Sequence 571840,
	303	8	3.8	604	6	US-10-149-759-29	Sequence 29, Appl
	304	8	3.8	605	8	US-10-357-930-46015	Sequence 46015, A
	305	8	3.8	609	4	US-09-925-065A-630748	Sequence 630748,
c	306	8	3.8	610	6	US-10-027-632-246617	Sequence 246617,
c	307	8	3.8	610	6	US-10-027-632-246617	Sequence 246617,
c	308	8	3.8	611	4	US-09-925-065A-893900	Sequence 893900,
	309	8	3.8	612	7	US-10-437-963-14181	Sequence 14181, A
	310	8	3.8	613	5	US-10-027-632-286122	Sequence 286122,
	311	8	3.8	613	6	US-10-027-632-286123	Sequence 286123,
	312	8	3.8	613	5	US-10-027-632-286124	Sequence 286124,
	313	8	3.8	613	6	US-10-027-632-286122	Sequence 286122,
	314	8	3.8	613	6	US-10-027-632-286123	Sequence 286123,
	315	8	3.8	613	6	US-10-027-632-286124	Sequence 286124,
	316	8	3.8	619	6	US-10-631-467-414	Sequence 414, App
c	317	8	3.8	621	9	US-09-925-065A-894636	Sequence 894636,
c	318	8	3.8	622	4	US-09-925-065A-464419	Sequence 464419,
	319	8	3.8	622	8	US-10-363-345A-37883	Sequence 37883, A
	320	8	3.8	622	8	US-10-363-345A-37884	Sequence 37884, A
c	321	8	3.8	622	9	US-10-363-483A-37883	Sequence 37883, A
	322	8	3.8	622	9	US-10-363-483A-37884	Sequence 37884, A
	323	8	3.8	623	3	US-09-954-456-1843	Sequence 1843, Ap
	324	8	3.8	623	4	US-09-925-065A-672678	Sequence 672678,
	325	8	3.8	623	9	US-09-925-065A-686260	Sequence 686260,
	326	8	3.8	623	9	US-10-240-425-1302	Sequence 1302, Ap
	327	8	3.8	623	7	US-10-843-641A-4870	Sequence 4870, Ap
c	328	8	3.8	624	9	US-09-925-065A-31084	Sequence 31084,
c	329	8	3.8	624	4	US-10-424-599-749727	Sequence 749727,
c	330	8	3.8	624	4	US-09-925-065A-822330	Sequence 822330,
	331	8	3.8	625	7	US-10-425-115-63687	Sequence 63687, A
	332	8	3.8	627	4	US-10-437-963-43053	Sequence 43053, A
c	333	8	3.8	630	6	US-09-796-692-8050	Sequence 8050, Ap
c	334	8	3.8	630	6	US-10-040-862-8050	Sequence 8050, Ap
c	335	8	3.8	630	6	US-10-057-475B-8050	Sequence 8050, Ap
c	336	8	3.8	630	6	US-10-154-884B-8050	Sequence 8050, Ap
	337	8	3.8	630	8	US-10-764-324-8050	Sequence 8050, Ap
c	338	8	3.8	632	7	US-10-424-599-63942	Sequence 63942, A
c	339	8	3.8	637	4	US-09-925-065A-883750	Sequence 883750,
	340	8	3.8	637	4	US-09-925-065A-910175	Sequence 910175,
c	341	8	3.8	637	5	US-10-027-632-212274	Sequence 212274,
c	342	8	3.8	637	5	US-10-027-632-212275	Sequence 212275,
	343	8	3.8	637	6	US-10-027-632-212276	Sequence 212276,
	344	8	3.8	637	5	US-10-027-632-223453	Sequence 223453,
c	345	8	3.8	637	5	US-10-027-632-223454	Sequence 223454,
	346	8	3.8	637	5	US-10-027-632-223455	Sequence 223455,
	347	8	3.8	637	6	US-10-027-632-212274	Sequence 212274,
c	348	8	3.8	637	6	US-10-027-632-212275	Sequence 212275,
	349	8	3.8	637	6	US-10-027-632-212276	Sequence 212276,
	350	8	3.8	637	6	US-10-027-632-223453	Sequence 223453,
	351	8	3.8	637	6	US-10-027-632-223454	Sequence 223454,
c	352	8	3.8	637	6	US-10-027-632-223455	Sequence 223455,
c	353	8	3.8	639	6	US-10-027-632-212274	Sequence 212274,
c	354	8	3.8	642	9	US-09-925-065A-709912	Sequence 709912,
	355	8	3.8	642	3	US-10-991-936-1609	Sequence 1609, Ap
	356	8	3.8	645	7	US-10-978-245-1609	Sequence 1609, Ap
	357	8	3.8	653	7	US-10-276-774-770	Sequence 770, App
c	358	8	3.8	658	5	US-10-027-632-77428	Sequence 77428, A
c	359	8	3.8	658	5	US-10-027-632-77428	Sequence 77428, A
c	360	8	3.8	658	6	US-10-027-632-320489	Sequence 320489,
c	361	8	3.8	661	7	US-10-027-632-320489	Sequence 320489,
	362	8	3.8	662	4	US-10-424-599-11432	Sequence 11432, A
c	363	8	3.8	663	7	US-09-925-065A-441386	Sequence 441386,
	364	8	3.8	663	4	US-10-437-963-83352	Sequence 83352, A
	365	8	3.8	664	4	US-09-925-065A-4577	Sequence 4577, Ap
	366	8	3.8	664	5	US-09-925-065A-4578	Sequence 4578, Ap
	367	8	3.8	664	6	US-09-925-065A-4579	Sequence 4579, Ap
c	368	8	3.8	665	7	US-10-437-963-36289	Sequence 36289, A
c	369	8	3.8	695	5	US-10-027-632-137416	Sequence 137416,
c	370	8	3.8	695	7	US-10-027-632-137416	Sequence 137416,
c	371	8	3.8	701	7	US-10-437-963-12876	Sequence 12876, A
	372	8	3.8	703	8	US-10-425-115-32048	Sequence 32048, A

Left column:

```
     373 8 9 3.8  US-10-971-986-1       Sequence 1, Appli
     703 8 9 3.8  US-10-971-339-1       Sequence 1, Appli
     710 8 7 3.8  US-10-467-534-91      Sequence 91, Appl
     712 8 7 3.8  US-10-425-115-154884  Sequence 154884,
     726 8 7 3.8  US-10-437-963-34580   Sequence 34580, A
     732 8 5 3.8  US-10-027-632-131017  Sequence 131017,
     732 8 5 3.8  US-10-027-632-150209  Sequence 150209,
     732 8 6 3.8  US-10-027-632-131017  Sequence 131017,
     732 8 6 3.8  US-10-027-632-150209  Sequence 150209,
     737 8 7 3.8  US-10-437-963-62647   Sequence 62647, A
     739 8 7 3.8  US-10-425-114-31132   Sequence 31132, A
     740 8 7 3.8  US-10-425-115-33448   Sequence 33448, A
     740 8 8 3.8  US-10-363-345A-27679  Sequence 27679, A
     740 8 9 3.8  US-10-363-345A-27680  Sequence 27680, A
     740 8 9 3.8  US-10-363-483A-27679  Sequence 27679, A
     743 8 8 3.8  US-10-363-483A-27680  Sequence 27680, A
     744 8 7 3.8  US-10-425-115-61902   Sequence 61902, A
     744 8 7 3.8  US-10-437-963-11272   Sequence 11272, A
     752 8 5 3.8  US-10-767-701-8708    Sequence 8708, Ap
     752 8 6 3.8  US-10-027-632-145875  Sequence 145875,
     760 8 7 3.8  US-10-027-632-145875  Sequence 145875,
     763 8 7 3.8  US-10-425-115-150094  Sequence 150094,
     768 8 3 3.8  US-10-425-114-838     Sequence 838, App
     770 8 6 3.8  US-09-814-355-20200   Sequence 20200, A
     770 8 6 3.8  US-10-027-632-150921  Sequence 150921,
     774 8 7 3.8  US-10-027-632-150921  Sequence 150921,
     776 8 8 3.8  US-10-437-963-98820   Sequence 98820, A
     789 8 5 3.8  US-10-425-115-65838   Sequence 65838, A
     789 8 6 3.8  US-10-027-632-156619  Sequence 156619,
     796 8 7 3.8  US-10-027-632-156619  Sequence 156619,
     815 8 6 3.8  US-10-425-115-144360  Sequence 144360,
     818 8 7 3.8  US-10-425-114-34687   Sequence 34687, A
     821 8 6 3.8  US-10-257-826A-259    Sequence 259, App
     830 8 3 3.8  US-10-264-049-2104    Sequence 2104, Ap
     830 8 3 3.8  US-09-870-759-35      Sequence 35, Appl
     830 8 3 3.8  US-09-751-708A-35     Sequence 35, Appl
     830 8 9 3.8  US-10-428-817A-31     Sequence 31, Appl
     842 8 8 3.8  US-10-937-758A-35     Sequence 35, Appl
     842 8 8 3.8  US-10-363-345A-21177  Sequence 21177, A
     842 8 9 3.8  US-10-363-345A-21178  Sequence 21178, A
     842 8 9 3.8  US-10-363-483A-21177  Sequence 21177, A
     850 8 8 3.8  US-10-363-483A-21178  Sequence 21178, A
     859 8 8 3.8  US-10-425-114-217512  Sequence 137512,
     869 8 8 3.8  US-10-723-860-5642    Sequence 5642, Ap
     875 8 8 3.8  US-10-425-115-101396  Sequence 101396,
     894 8 3 3.8  US-09-813-289-19      Sequence 19, Appl
     894 8 6 3.8  US-10-259-165-722     Sequence 722, App
     895 8 5 3.8  US-10-425-115-20341   Sequence 20341, A
     897 8 7 3.8  US-10-425-114-15404   Sequence 15404, A
     897 8 7 3.8  US-10-437-963-81545   Sequence 81545, A
     905 8 6 3.8  US-10-027-632-171324  Sequence 171324,
     905 8 6 3.8  US-10-027-632-171324  Sequence 171324,
     909 8 8 3.8  US-10-335-053-198     Sequence 198, App
     912 8 7 3.8  US-10-369-493-24216   Sequence 24216, A
     912 8 7 3.8  US-10-437-963-43055   Sequence 43055, A
     912 8 8 3.8  US-10-425-115-50044   Sequence 50044, A
     915 8 7 3.8  US-10-437-963-59367   Sequence 59367, A
     915 8 7 3.8  US-10-425-115-70832   Sequence 70832, A
     916 8 7 3.8  US-10-425-114-33400   Sequence 33400, A
     926 8 7 3.8  US-10-425-114-27501   Sequence 27501, A
     945 8 5 3.8  US-10-027-632-170287  Sequence 170287,
     949 8 8 3.8  US-10-027-632-170287  Sequence 170287,
     949 8 5 3.8  US-10-425-115-37992   Sequence 37992, A
     950 8 8 3.8  US-10-243-552-742     Sequence 742, App
     957 8 6 3.8  US-10-425-114-4810    Sequence 4810, Ap
     963 8 5 3.8  US-10-001-857-43      Sequence 43, Appl
     965 8 7 3.8  US-10-437-963-35130   Sequence 35130, A
     971 8 8 3.8  US-10-437-963-94784   Sequence 94784, A
     971 8 7 3.8  US-10-424-599-111249  Sequence 111249,
     975 8 7 3.8  US-10-369-493-111249  Sequence 111249,
     996 8 5 3.8  US-10-027-632-323591  Sequence 323591,
     996 8 6 3.8  US-10-027-632-323591  Sequence 323591,
     998 8 5 3.8  US-10-027-632-323591  Sequence 323591,
     998 8 6 3.8  US-10-027-632-323591  Sequence 323591,
```

Right column:

```
c 446 8 3.8 1006  US-10-424-599-124368   Sequence 124368,
  447 8 3.8 1007  US-10-425-115-40817    Sequence 40817, A
c 448 6 3.8 1011  US-10-101-510-27       Sequence 27, Appl
c 449 8 3.8 1014  US-10-425-115-177080   Sequence 177080,
c 450 8 3.8 1020  US-10-921-707-29       Sequence 29, Appl
c 451 8 3.8 1032  US-10-425-114-34810    Sequence 34810, A
c 452 4 3.8 1038  US-09-925-065A-63837   Sequence 63837, A
c 453 8 3.8 1038  US-09-925-065A-63838   Sequence 63838, A
c 454 3 3.8 1038  US-09-925-065A-63839   Sequence 63839, A
c 455 3 3.8 1044  US-09-969-708-167      Sequence 167, App
c 456 3 3.8 1044  US-09-880-107-2241     Sequence 2241, Ap
c 457 6 3.8 1044  US-09-887-367C-702     Sequence 702, App
c 458 9 3.8 1044  US-10-341-434-148      Sequence 148, App
c 459 8 3.8 1044  US-10-843-641A-702     Sequence 702, App
c 460 9 3.8 1044  US-10-844-641A-7638    Sequence 7638, Ap
c 461 9 3.8 1044  US-10-956-157-1039     Sequence 1039, Ap
c 462 9 3.8 1044  US-10-956-157-6274     Sequence 6274, Ap
c 463 3 3.8 1053  US-10-425-115-181896   Sequence 181896,
c 464 8 3.8 1054  US-09-951-938-10       Sequence 10, Appl
c 465 7 3.8 1054  US-10-447-839A-10      Sequence 10, Appl
c 466 4 3.8 1054  US-10-778-859-10       Sequence 10, Appl
c 467 8 3.8 1056  US-09-925-065A-576     Sequence 576, App
c 468 4 3.8 1056  US-09-925-065A-577     Sequence 577, App
c 469 4 3.8 1056  US-09-925-065A-578     Sequence 578, App
c 470 4 3.8 1056  US-09-925-065A-579     Sequence 579, App
c 471 7 3.8 1062  US-10-437-963-34157    Sequence 34157, A
c 472 9 3.8 1065  US-10-655-047-7080     Sequence 7080, Ap
c 473 9 3.8 1074  US-10-450-763-9792     Sequence 9792, Ap
c 474 9 3.8 1083  US-10-450-763-10479    Sequence 10479, A
c 475 7 3.8 1087  US-09-770-445-117      Sequence 117, App
c 476 7 3.8 1090  US-10-437-963-79952    Sequence 79952, A
c 477 7 3.8 1099  US-10-282-122A-23961   Sequence 23961, A
c 478 8 3.8 1103  US-10-425-115-23695    Sequence 23695, A
c 479 7 3.8 1128  US-10-425-114-32631    Sequence 32631, A
c 480 7 3.8 1137  US-10-437-963-90262    Sequence 90262, A
c 481 6 3.8 1137  US-10-369-493-45518    Sequence 45518, A
c 482 7 3.8 1140  US-10-437-963-81542    Sequence 81542, A
c 483 9 3.8 1140  US-10-945-678-31       Sequence 31, Appl
c 484 7 3.8 1140  US-10-450-763-10530    Sequence 10530, A
c 485 7 3.8 1144  US-10-424-599-41667    Sequence 41667, A
c 486 3 3.8 1146  US-09-908-671-1        Sequence 1, Appli
c 487 8 3.8 1161  US-09-974-300-2720     Sequence 2720, Ap
c 488 10 3.8 1163 US-11-037-657-1        Sequence 1, Appli
c 489 7 3.8 1173  US-10-437-963-91611    Sequence 91611, A
c 490 9 3.8 1186  US-10-956-157-4796     Sequence 4796, App
c 491 9 3.8 1186  US-10-956-157-10031    Sequence 10031, A
c 492 4 3.8 1187  US-09-925-065A-279504  Sequence 279504,
c 493 6 3.8 1187  US-10-510-548          Sequence 548, App
c 494 3 3.8 1200  US-09-960-706-1050     Sequence 1050, A
c 495 3 3.8 1200  US-09-873-319-697      Sequence 697, App
c 496 7 3.8 1200  US-10-425-114-27522    Sequence 27522, A
c 497 7 3.8 1206  US-10-437-963-81103    Sequence 81103, A
c 498 6 3.8 1209  US-10-166-653-17       Sequence 17, Appl
c 499 10 3.8 1209 US-11-116-192-17       Sequence 17, Appl
c 500 8 3.8 1214  US-10-425-115-101397   Sequence 101397,
```

ALIGNMENTS

RESULT 1
US-09-989-890-105
; Sequence 105, Application US/09989890
; Publication No. US20040166105A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Pluta, Jason
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0287
; CURRENT APPLICATION NUMBER: US/09/989,890

; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,509
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 105
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-989-890-105

Alignment Scores:
Pred. No.:
Score: 8.51e-192
Percent Similarity: 212.00
Best Local Similarity: 100.0%
Query Match: 100.0%
DB:

Length: 1713
Matches: 212
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 3

US-09-989-890-238 (1-212) x US-09-989-890-105 (1-1713)

```
Qy      1 SerProHisGlnAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
Db    104 AGCCCCCACCAAGCGCCGCACCCGTAGACCAGACCCCAGGACCCTGGCCACCATGGGC      163

Qy     21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAla       40
Db    164 CAGAGAGCATTACCTTCATCTCTGGCTCTTGCTCTGCTGAGCCGGCCCTTGAGTCCCCACCTGCT  223

Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla      60
Db    224 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGCTGCCTTCTTCGCTTCCGCCGCT     283

Qy     61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu      80
Db    284 GCCGGGATTGCCTCCAGCCGCTGTGTGGGGGATGCAGCCCCCTGCCTGTCTA        343

Qy     81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys     100
Db    344 CTGAGGACTCCACTGAGGGGACTGCTGAGCCAACTGGGCCAAGGAGCACAATGGAGTGC       403

Qy    101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGlnPro             120
Db    404 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAGCAAGTCAACCA       463

Qy    121 TrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProThr     140
Db    464 TGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGCTCCCTGTGTATCCCTACC       523

Qy    141 ArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIle     160
Db    524 CGAGGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGAT       583

Qy    161 ProHisProCysAspThrAlaCysProAlaProLeuProValLeuValAlaProArg        180
Db    584 CCCCACCCATGCGACACAGCCTGCCCAGCACCTTTGCCGTTCTCGTGGGTCCGAGG         643

Qy    181 SerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCys     200
Db    644 AGTACTATTCTTTCCATGAGTCGGACCTGGACCTGGCAGTGGCGGCTCCATGT            703

Qy    201 ArgAlaGluLysLeuMetCysSerSerArgSer   212
Db    704 CGAGCCGAGAAATTGATGTGCCTCATCTTCAAGAAGC   739
```

RESULT 2
US-10-104-047-799
; Sequence 799, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105

; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 799
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-799

Alignment Scores:
Pred. No.:
Score: 2.46e-99
Percent Similarity: 115.00
Best Local Similarity: 99.1%
Query Match: 99.1%
DB: 54.2%

Length: 1785
Matches: 211
Conservative: 0
Mismatches: 1
Indels: 2
Gaps: 0

US-09-989-890-238 (1-212) x US-10-104-047-799 (1-1785)

```
Qy      1 SerProHisGlnAlaAlaAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
Db    193 AGCCCCCACCAAGCGCCGCACCCGTAGACCAGACCCCAGGACCCTGGCCACCATGGGC      252

Qy     21 GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAla       40
Db    253 CAGAGAGCATTACCTTCATCTCTGGCTCTTGCTCTGCTGAGCCGGCCCTTGAGTCCCCACCTGCT  312

Qy     41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla      60
Db    313 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGCTGCCTTCTTCGCTTCCGCCGCT     372

Qy     61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu      80
Db    373 GCCGGGATTGCCTCCAGCCGCTGTGTGGGGGATGCAGCCCCCTGCCTGTCTA        432

Qy     81 LeuArgThrProLeuArgGlyLeuLeuLysProThrGlyProArgSerThrMetGluCys     100
Db    433 CTGAGGACTCCACTGAGGGGACTGCTGAGCCAACTGGGCCAAGGAGCACAATGGAGTGC       492

Qy    101 ProProAlaLeuIleValHisProProAlaGlyGlyMetAlaSerGly-SerSerGlnPr     120
Db    493 CCCCCAGCCCTGATCGTGCACCCCCCAGCCGGCGGGGATGGCCAGCAG-CTCAAGTCAACC     551

Qy    120 oTrpAlaAlaAlaSerAlaThrProMetLeuSerSerLysAlaSerLeuCysIleProTh     140
Db    552 ATGGGCAGCAGCTTCAGCTACCCCGATGTTAAGCTCAAAGCTCCCTGTGTATCCCTAC       611

Qy    140 rArgGlyProProGlnProLeuMetArgThrProAlaAlaAlaArgSerHisTrpProIl     160
Db    612 CCGAGGGCCACCTCCCCAGCCCCTGATGCGGACTCCTGCTGCAAGGAGCCACTGGCCGAT      671

Qy    160 eProHisProCysAspThrAlaCysProAlaProLeuProValLeuValAlaProAr       180
Db    672 CCCCACCCATGCGACACAGCCTGCCCAGCACCTTTGCCGTTCTCGTGGGTCCGAG          731

Qy    180 gSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCy     200
Db    732 GAGTACTATTCTTTCCATGAGTCGGACCTGGACCTGGCAGTGGCGGCTCCATG            791

Qy    200 sArgAlaGluLysLeuMetCysSerSerArgSer   212
Db    792 TCGAGCCGAGAAATTGATGTGCCTCATCTTCAAGAAGC   828
```

RESULT 3
US-10-779-543-12051
; Sequence 12051, Application US/10779543
; Publication No. US20050227917A1
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED

; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 12051
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 401
; OTHER INFORMATION: n = A,T,C or G
US-10-779-543-12051

Alignment Scores:
Pred. No.:   1.88e-67   Length: 427
Score: 81.00   Matches: 81
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 38.2%   Indels: 0
DB: Gaps: 9

US-09-989-890-238 (1-212) x US-10-779-543-12051 (1-427)

Qy   1 SerProHisGlnAlaAlaAlaProValAspProArgThrLeuAlaThrMetGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183 AGCCCCCACCAAGCGCCGCACCCGTAGACCAGACCCCGCCACCATGGGC   242

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGCCCCACCTGCT   302

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303 GCCTGCTCTGGCGACCCTGGGTGTGGTGGCCGGCGGTGGGCGGGGCTGCCCTGCGCCGCT 362

Qy   61 AlaGlyIleAlaSerAlaValGluProValCysGlyAspAlaProAlaCysLeu 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 GCCGGATTGCCTCCAGCCGTGTGCGAGCCTGTGCGGNGATGCAGCCCCTGCTGTCTA   422

Qy   81 Leu 81
          |||
Db   423 CTG 425

RESULT 4
US-10-074-475-124
; Publication No. US20030092898A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana

<table>
<tr><td>;</td><td>APPLICANT:</td><td>Macina, Roberto</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Hu, Ping</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Recipon, Herve</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Karra, Kalpana</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Cafferkey, Robert</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Sun, Yongming</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Liu, Chenghua</td></tr>
</table>

; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0313
; CURRENT APPLICATION NUMBER: US/10/074,475
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: 60/268,292
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 295
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 124
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-074-475-124

Alignment Scores:
Pred. No.:   7.34e-56   Length: 654
Score: 69.00   Matches: 105
Percent Similarity: 98.1%   Conservative: 0
Best Local Similarity: 98.1%   Mismatches: 1
Query Match: 32.5%   Indels: 2
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-074-475-124 (1-654)

Qy   1 SerProHisGlnAlaAlaAlaProValAspProArgThrLeuAlaThrMetGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16 AGCCCCCACCAAGCGCCGCACCCGTAGACCAGACCCCAAGGACCCCTGGCCACCATGGGC 75

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 CAGAGAGCATTACCTTCATCTCTGGCTCTGCTGAGCCGCCGGCCCTTGAGTCCCCACCTGCT 135

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   136 GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGGCCGGCGGTGGGCGGGGCTGCCCTGGCT 195

Qy   61 AlaGlyIleAlaSerSerAlaValGluPro--ValCysGlyAspAlaProAlaCysLe 80
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   196 GCCGGGATTGCCTCCCAGCCGTGTGGAGGC-CGTGTGCGGGGGGATGCAGCCCCTGCCTGTCT 254

Qy   80 uLeuArgThrProLeuArgGlyLeuLeuLysProThrGlySerThrMetGluCy 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   255 ACTGAGGACTCCCCTGAGGGGACTGCTGAGCCAACTGGTCCAAGGAGCACAATGGAGTG 314

Qy   100 sProProAlaLeuIleVal 106
          |||||||||||||||||||
Db   315 CCCCCCAGCCCTGATCGTG 333

RESULT 5
US-09-989-920-56
; Sequence 56, Application US/09989920
; Patent No. US2002017295781
; GENERAL INFORMATION:

<table>
<tr><td>;</td><td>APPLICANT:</td><td>Macina, Roberto</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Recipon, Herve</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Chen, Sei-Yu</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Sun, Yongming</td></tr>
<tr><td>;</td><td>APPLICANT:</td><td>Liu, Chenghua</td></tr>
</table>

; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284

```
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 56
;    LENGTH: 1977
;    TYPE: DNA
;    ORGANISM: Homo sapien
US-09-989-920-56

Alignment Scores:
Pred. NO.:                    6.29e-46        Length:              1977
Score:                        59.00           Matches:             59
Percent Similarity:           100.0%          Conservative:        0
Best Local Similarity:        100.0%          Mismatches:          0
Query Match:                  27.8%           Indels:              0
DB:                           3               Gaps:                0

US-09-989-890-238 (1-212) x US-09-989-920-56 (1-1977)

Qy      1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357  AGCCCCCACCAAGCCCGCACCCGTAGACCAGCAGCACCCCGGCCACCATGGGC  416

Qy     21  GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAla  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417  CAGAGAGCATTACCTTCATCTCGCTCTCTGAGCCGGCCCTTGAGTCCCCACCTGCT  476

Qy     41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla  59
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477  GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCC  533

RESULT 6
US-09-989-920-117
; Sequence 117, Application US/09989920
; Patent No. US2002017295741
; GENERAL INFORMATION:
;    APPLICANT: Macina, Roberto
;    APPLICANT: Recipon, Herve
;    APPLICANT: Chen, Sei-Yu
;    APPLICANT: Sun, Yongming
;    APPLICANT: Liu, Chenghua
;    TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
;    FILE REFERENCE: DEX-0291
;    CURRENT APPLICATION NUMBER: US/09/989,920
;    CURRENT FILING DATE: 2001-11-21
;    PRIOR APPLICATION NUMBER: 60/252,500
;    PRIOR FILING DATE: 2000-11-22
;    NUMBER OF SEQ ID NOS: 284
;    SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 117
;    LENGTH: 1977
;    TYPE: DNA
;    ORGANISM: Homo sapien
US-09-989-920-117

Alignment Scores:
Pred. NO.:                    6.29e-46        Length:              1977
Score:                        59.00           Matches:             59
Percent Similarity:           100.0%          Conservative:        0
Best Local Similarity:        100.0%          Mismatches:          0
Query Match:                  27.8%           Indels:              0
DB:                           3               Gaps:                0

US-09-989-890-238 (1-212) x US-09-989-920-117 (1-1977)

Qy      1  SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly  20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357  AGCCCCCACCAAGCCCGCACCCGTAGACCAGCAGCACCCCGGCCACCATGGGC  416

Qy     21  GlnArgAlaLeuProSerLeuAlaLeuLeuSerArgProLeuSerProProProAla  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417  CAGAGAGCATTACCTTCATCTCGCTCTCTGAGCCGGCCCTTGAGTCCCCACCTGCT  476

Qy     41  AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAla  59
           |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      477  GCCTGCTCTGGCGACCCTGGGTGTGGGAGTGGTGCCGGGCTGCCTTCTGCTTCCGCC  533

RESULT 7
US-10-741-601-5611
; Sequence 5611, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;    APPLICANT: CARGILL, Michele et al.
;    TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;    TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
;    FILE REFERENCE: CL001500
;    CURRENT APPLICATION NUMBER: US/10/741,601
;    CURRENT FILING DATE: 2003-12-22
;    NUMBER OF SEQ ID NOS: 26415
;    SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 5611
;    LENGTH: 253861
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-741-601-5611

Alignment Scores:
Pred. No.:                                    Length:              253861
Score:                        205             Matches:             11
Percent Similarity:           100.0%          Conservative:        0
Best Local Similarity:        100.0%          Mismatches:          0
Query Match:                  5.2%            Indels:              0
DB:                           7               Gaps:                0

US-09-989-890-238 (1-212) x US-10-741-601-5611 (1-253861)

Qy     24  LeuProSerSerAlaAlaLeuSerArgPro  34
           ||||||||||||||||||||||||||||||
Db 237045  CTTCCATCCAGCCTCGCTCTTCCTGAGCCGGCCCC  237077

RESULT 8
US-10-087-192-2002
; Sequence 2002, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;    APPLICANT: Morris, David W.
;    APPLICANT: Engelhard, Eric K.
;    TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION: CANCER
;    FILE REFERENCE: 52945200D122
;    CURRENT APPLICATION NUMBER: US/10/087,192
;    CURRENT FILING DATE: 2002-03-01
;    PRIOR APPLICATION NUMBER: US 09/747,377
;    PRIOR FILING DATE: 2000-12-22
;    PRIOR APPLICATION NUMBER: US 09/798,586
;    PRIOR FILING DATE: 2001-03-02
;    NUMBER OF SEQ ID NOS: 2059
;    SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 2002
;    LENGTH: 261817
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-087-192-2002

Alignment Scores:
Pred. No.:                                    Length:              261817
Score:                        210             Matches:             11
Percent Similarity:           100.0%          Conservative:        0
Best Local Similarity:        100.0%          Mismatches:          0
Query Match:                  5.2%            Indels:              0
DB:                           5               Gaps:                0

US-09-989-890-238 (1-212) x US-10-087-192-2002 (1-261817)

Qy     24  LeuProSerSerAlaAlaLeuSerArgPro  34
           ||||||||||||||||||||||||||||||
Db 241066  CTTCCATCCAGCCTCGCTCTTCCTGAGCCGGCCCC  241098
```

RESULT 9
US-10-027-632-143483
; Sequence 143483, Application US2002019837141
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143483
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-143483

Alignment Scores:
Pred. No.:   10.3   Length: 523
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-027-632-143483 (1-523)

Qy   26 SerSerLeuAlaLeuSerArgProLeu 35
               ||||||||||||||||||||||||||||
Db   174 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 203

RESULT 10
US-10-027-632-143484
; Sequence 143484, Application US2002019837141
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143484
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-143484

Alignment Scores:
Pred. No.:   10.3   Length: 523
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-027-632-143484 (1-523)

Qy   26 SerSerLeuAlaLeuSerArgProLeu 35
               ||||||||||||||||||||||||||||
Db   174 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 203

RESULT 11
US-10-027-632-143483
; Sequence 143483, Application US2003020407549
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143483
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-143483

Alignment Scores:
Pred. No.:   10.3   Length: 523
Score: 10.00   Matches: 10
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.7%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-027-632-143483 (1-523)

Qy   26 SerSerLeuAlaLeuSerArgProLeu 35
               ||||||||||||||||||||||||||||
Db   174 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 203

RESULT 12
US-10-027-632-143484
; Sequence 143484, Application US2003020407549
; Publication No. US20030204075A9
; GENERAL INFORMATION:

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143484
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-143484

Alignment Scores:
Pred. No.:               10.3      Length:            523
Score:                   10.00     Matches:           10
Percent Similarity:      100.0%    Conservative:      0
Best Local Similarity:   100.0%    Mismatches:        0
Query Match:             4.7%      Indels:            0
DB:                      6         Gaps:              0

US-09-989-890-238 (1-212) x US-10-027-632-143484 (1-523)

Qy      26 SerSerLeuAlaLeuLeuSerArgProLeu 35
              ||||||||||||||||||||||||||||
Db     174 TCCTCACTGGCCCTTCTATCCAGGCCCCTT 203


RESULT 13
US-10-156-761-4893/c
; Sequence 4893, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4893
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(861)
US-10-156-761-4893

Alignment Scores:
Pred. No.:               15.7      Length:            861
Score:                   10.00     Matches:           10
Percent Similarity:      100.0%    Conservative:      0
Best Local Similarity:   100.0%    Mismatches:        0
Query Match:             4.7%      Indels:            0
DB:                      6         Gaps:              0

US-09-989-890-238 (1-212) x US-10-156-761-4893 (1-861)

Qy      51 GlyAlaGlyLeuProSerAlaSerAlaAla 60
              ||||||||||||||||||||||||||||
Db     668 GGAGCGGGCTTGCCCTCCGGCCTCGGCGGCC 639


RESULT 14
US-10-007-926A-134
; Sequence 134, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: FERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; TITLE OF INVENTION: USING ARRAYS OF CANDIDATE GENES
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 134
; LENGTH: 5723
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: forkhead box o1a (rhabdomyosarcoma) (FOXO1A)
; OTHER INFORMATION: gene.
US-10-007-926A-134

Alignment Scores:
Pred. No.:               76.6      Length:            5723
Score:                   10.00     Matches:           10
Percent Similarity:      100.0%    Conservative:      0
Best Local Similarity:   100.0%    Mismatches:        0
Query Match:             4.7%      Indels:            0
DB:                      6         Gaps:              0

US-09-989-890-238 (1-212) x US-10-007-926A-134 (1-5723)

Qy      52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
              ||||||||||||||||||||||||||||
Db     551 GCGGGCCTGCCCTGCCCTCGGCCTGCCGCT 580


RESULT 15
US-10-341-434-52
; Sequence 52, Application US/10341434
; Publication No. US2003021583A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
```

```
; LENGTH: 5723
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (386)..(2350)
; OTHER INFORMATION:
US-10-341-434-52

Alignment Scores:
Pred. No.:               76.6          Length:        5723
Score:                   10.00         Matches:       10
Percent Similarity:      100.0%        Conservative:  0
Best Local Similarity:   100.0%        Mismatches:    0
Query Match:             4.7%          Indels:        0
DB:                      6             Gaps:          0

US-09-989-890-238 (1-212) x US-10-341-434-52 (1-5723)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   551 GCGGGCCTGCCCTCGGCTGCCGCT 580

RESULT 16
US-10-671-074-4
; Sequence 4, Application US/10671074
; Publication No. US20040097459A1
; GENERAL INFORMATION:
; APPLICANT: Dobie, Kenneth W.
; APPLICANT: Bhanot, Sanjay
; APPLICANT: Veniant-Ellison, Murielle
; APPLICANT: Lindberg, Richard A.
; APPLICANT: Shutter, John R.
; TITLE OF INVENTION: MODULATION OF FORKHEAD BOX O1A EXPRESSION
; FILE REFERENCE: AMGN0001-101
; CURRENT APPLICATION NUMBER: US/10/671,074
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US 10/260,203
; PRIOR FILING DATE: 2002-09-26
; NUMBER OF SEQ ID NOS: 176
; SEQ ID NO 4
; LENGTH: 5723
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (386)..(2353)
US-10-671-074-4

Alignment Scores:
Pred. No.:               76.6          Length:        5723
Score:                   10.00         Matches:       10
Percent Similarity:      100.0%        Conservative:  0
Best Local Similarity:   100.0%        Mismatches:    0
Query Match:             4.7%          Indels:        0
DB:                      7             Gaps:          0

US-09-989-890-238 (1-212) x US-10-671-074-4 (1-5723)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   551 GCGGGCCTGCCCTCGGCTGCCGCT 580

RESULT 17
US-10-956-157-760
; Sequence 760, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
```

```
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 760
; LENGTH: 5723
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-760

Alignment Scores:
Pred. No.:               76.6          Length:        5723
Score:                   10.00         Matches:       10
Percent Similarity:      100.0%        Conservative:  0
Best Local Similarity:   100.0%        Mismatches:    0
Query Match:             4.7%          Indels:        0
DB:                      9             Gaps:          0

US-09-989-890-238 (1-212) x US-10-956-157-760 (1-5723)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   551 GCGGGCCTGCCCTCGGCTGCCGCT 580

RESULT 18
US-10-450-763-20857
; Sequence 20857, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 20857
; LENGTH: 5833
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (157)..(585)
; OTHER INFORMATION: 31% homologous to Homo sapiens fork head domain
; OTHER INFORMATION: protein, accession number U02310,Smith-Waterman Score=53.
US-10-450-763-20857

Alignment Scores:
Pred. No.:               77.9          Length:        5833
Score:                   10.00         Matches:       10
Percent Similarity:      100.0%        Conservative:  0
Best Local Similarity:   100.0%        Mismatches:    0
Query Match:             4.7%          Indels:        0
DB:                      9             Gaps:          0

US-09-989-890-238 (1-212) x US-10-450-763-20857 (1-5833)

Qy    52 AlaGlyLeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||||||||
Db   552 GCGGGCCTGCCCTCGGCTGCCGCT 581

RESULT 19
US-10-155-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
```

```
;   APPLICANT:  OMURA, SATOSHI
;   APPLICANT:  IKEDA, HARUO
;   APPLICANT:  ISHIKAWA, JUN
;   APPLICANT:  HORIKAWA, HIROSHI
;   APPLICANT:  SHIBA, TADAYOSHI
;   APPLICANT:  SAKAKI, YOSHIYUKI
;   APPLICANT:  HATTORI, MASAHIRA
;   TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;   FILE REFERENCE: 249-262
;   CURRENT APPLICATION NUMBER: US/10/156,761
;   CURRENT FILING DATE:  2002-05-29
;   PRIOR APPLICATION NUMBER: JP 2001-204089
;   PRIOR FILING DATE: 2001-05-30
;   PRIOR APPLICATION NUMBER: JP 2001-272697
;   PRIOR FILING DATE: 2001-08-02
;   NUMBER OF SEQ ID NOS: 15109
;   SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.:               3.66e+04            Length:         9025608
Score:                   10.00               Matches:        10
Percent Similarity:      100.0%              Conservative:   0
Best Local Similarity:   100.0%              Mismatches:     0
Query Match:             4.7%                Indels:         0
DB:                      6                   Gaps:           0

US-09-989-890-238 (1-212) x US-10-156-761-1 (1-9025608)

Qy       51 GlyAlaGlyLeuProSerAlaSerAlaAla 60
            ||||||||||||||||||||||||||||||
Db  5965745 GGAGCGGGGCTTGCCCTCGGCCTCGGCGGCC 5965716

RESULT 20
US-10-719-993-6100/c
; Sequence 6100, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;   APPLICANT: CARGILL, Michele et al.
;   TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;   TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001496
;   CURRENT APPLICATION NUMBER: US/10/719,993
;   CURRENT FILING DATE:  2003-11-24
;   NUMBER OF SEQ ID NOS: 55342
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 6100
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6100

Alignment Scores:
Pred. No.:               41.6                Length:         201
Score:                   9.00                Matches:        9
Percent Similarity:      100.0%              Conservative:   0
Best Local Similarity:   100.0%              Mismatches:     0
Query Match:             4.2%                Indels:         0
DB:                      8                   Gaps:           0

US-09-989-890-238 (1-212) x US-10-719-993-6100 (1-201)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db   172 GCTCTTCTTTCACGGCCCACTTTCCCCA 146
```

```
RESULT 21
US-10-719-993-6190/c
; Sequence 6190, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;   APPLICANT: CARGILL, Michele et al.
;   TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;   TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001496
;   CURRENT APPLICATION NUMBER: US/10/719,993
;   CURRENT FILING DATE:  2003-11-24
;   NUMBER OF SEQ ID NOS: 55342
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 6190
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6190

Alignment Scores:
Pred. No.:               41.6                Length:         201
Score:                   9.00                Matches:        9
Percent Similarity:      100.0%              Conservative:   0
Best Local Similarity:   100.0%              Mismatches:     0
Query Match:             4.2%                Indels:         0
DB:                      8                   Gaps:           0

US-09-989-890-238 (1-212) x US-10-719-993-6190 (1-201)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db   172 GCTCTTCTTTCACGGCCCACTTTCCCCA 146

RESULT 22
US-10-719-993-6280/c
; Sequence 6280, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;   APPLICANT: CARGILL, Michele et al.
;   TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;   TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001496
;   CURRENT APPLICATION NUMBER: US/10/719,993
;   CURRENT FILING DATE:  2003-11-24
;   NUMBER OF SEQ ID NOS: 55342
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 6280
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6280

Alignment Scores:
Pred. No.:               41.6                Length:         201
Score:                   9.00                Matches:        9
Percent Similarity:      100.0%              Conservative:   0
Best Local Similarity:   100.0%              Mismatches:     0
Query Match:             4.2%                Indels:         0
DB:                      8                   Gaps:           0

US-09-989-890-238 (1-212) x US-10-719-993-6280 (1-201)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db   172 GCTCTTCTTTCACGGCCCACTTTCCCCA 146

RESULT 23
US-10-719-993-6416/c
; Sequence 6416, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;   APPLICANT: CARGILL, Michele et al.
;   TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
;   TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001496
;   CURRENT APPLICATION NUMBER: US/10/719,993
;   CURRENT FILING DATE: 2003-11-24
;   NUMBER OF SEQ ID NOS: 55342
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 6416
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6416

Alignment Scores:
Pred. No.:
Score:                    41.6        Length:          201
Percent Similarity:      100.0%       Matches:           9
Best Local Similarity:   100.0%       Conservative:      0
Query Match:               4.2%       Mismatches:        0
DB:                          8        Indels:            0
                                      Gaps:              0

US-09-989-890-238 (1-212) x US-10-719-993-6416 (1-201)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db  172 GCTCTTCTTTCACGCCCACTTTCCCCA 146

RESULT 24
US-10-719-993-39711/c
;   Sequence 39711, Application US20040265849A1
;   Publication No. US20040265849A1
;   GENERAL INFORMATION:
;   APPLICANT: CARGILL, Michele et al.
;   TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;   TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001496
;   CURRENT APPLICATION NUMBER: US/10/719,993
;   CURRENT FILING DATE: 2003-11-24
;   NUMBER OF SEQ ID NOS: 55342
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 39711
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-39711

Alignment Scores:
Pred. No.:
Score:                    41.6        Length:          201
Percent Similarity:      100.0%       Matches:           9
Best Local Similarity:   100.0%       Conservative:      0
Query Match:               4.2%       Mismatches:        0
DB:                          8        Indels:            0
                                      Gaps:              0

US-09-989-890-238 (1-212) x US-10-719-993-39711 (1-201)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db  172 GCTCTTCTTTCACGCCCACTTTCCCCA 146

RESULT 25
US-10-243-552-741/c
;   Sequence 741, Application US20030224379A1
;   Publication No. US20030224379A1
;   GENERAL INFORMATION:
;   APPLICANT: Tang, Y., Tom
;   APPLICANT: Yang, Yonghong
;   APPLICANT: Wang, Zhiwei
;   APPLICANT: Weng, Gezhi
;   APPLICANT: Ma, Yunqing
;   TITLE OF INVENTION: Novel Nucleic Acids and
;   TITLE OF INVENTION: Polypeptides
;   FILE REFERENCE: 807A
;   CURRENT APPLICATION NUMBER: US/10/243,552
```

```
;   CURRENT FILING DATE: 2002-09-12
;   PRIOR APPLICATION NUMBER: US 60/322,511
;   PRIOR FILING DATE: 2001-09-13
;   PRIOR APPLICATION NUMBER: PCT/US00/35017
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: US 09/488,725
;   PRIOR FILING DATE: 2000-01-21
;   PRIOR APPLICATION NUMBER: US 09/552,317
;   PRIOR FILING DATE: 2000-04-25
;   PRIOR APPLICATION NUMBER: PCT/US01/02623
;   PRIOR FILING DATE: 2001-01-25
;   PRIOR APPLICATION NUMBER: US 09/491,404
;   PRIOR FILING DATE: 2000-01-25
;   PRIOR APPLICATION NUMBER: PCT/US01/03800
;   PRIOR FILING DATE: 2001-02-05
;   PRIOR APPLICATION NUMBER: US 09/496,914
;   PRIOR FILING DATE: 2000-02-03
;   PRIOR APPLICATION NUMBER: US 09/560,875
;   PRIOR FILING DATE: 2000-04-27
;   PRIOR APPLICATION NUMBER: PCT/US01/04927
;   PRIOR FILING DATE: 2001-02-26
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 998
;   SOFTWARE: pt_FL_genes Version 5.0
;   SEQ ID NO 741
;   LENGTH: 239
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-243-552-741

Alignment Scores:
Pred. No.:
Score:                    48.1        Length:          239
Percent Similarity:      100.0%       Matches:           9
Best Local Similarity:   100.0%       Conservative:      0
Query Match:               4.2%       Mismatches:        0
DB:                          6        Indels:            0
                                      Gaps:              0

US-09-989-890-238 (1-212) x US-10-243-552-741 (1-239)

Qy   54 LeuProSerAlaSerAlaAlaAlaGly 62
           |||||||||||||||||||||||||
Db  131 CTTCCTTCAGCGTCAGCAGCAGCTGGA 105

RESULT 26
US-10-450-763-10524/c
;   Sequence 10524, Application US/10450763
;   Publication No. US200050196754A1
;   GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc
;   TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;   FILE REFERENCE: 790CIP3/US
;   CURRENT APPLICATION NUMBER: US/10/450,763
;   CURRENT FILING DATE: 2003-06-11
;   PRIOR APPLICATION NUMBER: PCT/US01/08631
;   PRIOR FILING DATE: 2001-03-30
;   PRIOR APPLICATION NUMBER: 09/540,217
;   PRIOR FILING DATE: 2000-03-31
;   PRIOR APPLICATION NUMBER: 09/649,167
;   PRIOR FILING DATE: 2000-08-23
;   NUMBER OF SEQ ID NOS: 60736
;   SEQ ID NO 10524
;   SOFTWARE: Custom
;   LENGTH: 239
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIMILAR
;   LOCATION: (151)..(237)
;   OTHER INFORMATION: 75% homologous to Homo sapiens BB1=malignant cell expression-
;   OTHER INFORMATION: enhanced gene/tumor progression-enhanced gene,accession number
;   OTHER INFORMATION: S82470,Smith-Waterman Score=116.
US-10-450-763-10524
```

Alignment Scores:
Pred. No.:   Length: 239
Score: Matches: 9
Percent Similarity: 48.1   Conservative: 0
Best Local Similarity: 9.00   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 100.0%   Gaps: 0
                           4.2%
                           9

US-09-989-890-238 (1-212) x US-10-450-763-10524 (1-239)

Qy   54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
            ||||||||||||||||||||||||||||||
Db   131 CTTCCTTCAGCGTCAGCAGCAGCTCGA 105

RESULT 27
US-09-918-995-3937/c
; Sequence 3937, Application US/09918995
; Publication No. US2003007362A1
; GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;   TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;   FILE REFERENCE: 20411-756
;   CURRENT APPLICATION NUMBER: US/09/918,995
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US/09/235,076
;   PRIOR FILING DATE: 1999-01-20
;   NUMBER OF SEQ ID NOS: 38054
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3937
;   LENGTH: 496
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(496)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-3937

Alignment Scores:
Pred. No.:   Length: 496
Score: Matches: 9
Percent Similarity: 88.7   Conservative: 0
Best Local Similarity: 9.00   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 100.0%   Gaps: 0
                           4.2%
                           3

US-09-989-890-238 (1-212) x US-09-918-995-3937 (1-496)

Qy   25 ProSerSerLeuAlaLeuLeuSerArg 33
            |||||||||||||||||||||||||||
Db   159 CCGAGCTCCCTGGCTCTACTATCCAGG 133

RESULT 28
US-10-195-730-73/c
; Sequence 73, Application US/10195730
; Publication No. US200301444492A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et. al
;   TITLE OF INVENTION: 101 Human Secreted Proteins
;   FILE REFERENCE: PZ017P1
;   CURRENT APPLICATION NUMBER: US/10/195,730
;   CURRENT FILING DATE: 2002-07-16
;   PRIOR APPLICATION NUMBER: US/09/281,976
;   PRIOR FILING DATE: 1999-03-31
;   PRIOR APPLICATION NUMBER: 60/060,837
;   PRIOR FILING DATE: 1997-10-02
;   PRIOR APPLICATION NUMBER: 60/060,862
;   PRIOR FILING DATE: 1997-10-02
;   NUMBER OF SEQ ID NOS: 390
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73

;   LENGTH: 546
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-195-730-73

Alignment Scores:
Pred. No.:   Length: 546
Score: Matches: 9
Percent Similarity: 96.1   Conservative: 0
Best Local Similarity: 9.00   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 100.0%   Gaps: 0
                           4.2%
                           6

US-09-989-890-238 (1-212) x US-10-195-730-73 (1-546)

Qy   127 ThrProMetLeuSerSerLysAlaSer 135
            |||||||||||||||||||||||||||
Db   276 ACACCTATGCTCTCCTCCAAAGCTTCA 250

RESULT 29
US-10-799-747-73/c
; Sequence 73, Application US/10799747
; Publication No. US2004015725A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et. al
;   TITLE OF INVENTION: 101 Human Secreted Proteins
;   FILE REFERENCE: PZ017P1
;   CURRENT APPLICATION NUMBER: US/10/799,747
;   CURRENT FILING DATE: 2004-03-15
;   PRIOR APPLICATION NUMBER: US/10/195,730
;   PRIOR FILING DATE: 2002-07-16
;   PRIOR APPLICATION NUMBER: US/09/281,976
;   PRIOR FILING DATE: 1999-03-31
;   PRIOR APPLICATION NUMBER: 60/060,837
;   PRIOR FILING DATE: 1997-10-02
;   PRIOR APPLICATION NUMBER: 60/060,862
;   PRIOR FILING DATE: 1997-10-02
;   NUMBER OF SEQ ID NOS: 390
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73

;   LENGTH: 546
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-799-747-73

Alignment Scores:
Pred. No.:   Length: 546
Score: Matches: 9
Percent Similarity: 96.1   Conservative: 0
Best Local Similarity: 9.00   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 100.0%   Gaps: 0
                           4.2%
                           7

US-09-989-890-238 (1-212) x US-10-799-747-73 (1-546)

Qy   127 ThrProMetLeuSerSerLysAlaSer 135
            |||||||||||||||||||||||||||
Db   276 ACACCTATGCTCTCCTCCAAAGCTTCA 250

RESULT 30
US-10-979-183-73/c
; Sequence 73, Application US/10979183
; Publication No. US2005006994A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et. al
;   TITLE OF INVENTION: 101 Human Secreted Proteins
;   FILE REFERENCE: PZ017P1
;   CURRENT APPLICATION NUMBER: US/10/979,183
;   CURRENT FILING DATE: 2004-11-03
;   PRIOR APPLICATION NUMBER: US/10/799,747
;   PRIOR FILING DATE: 2004-03-15
;   PRIOR APPLICATION NUMBER: US/10/195,730
;   PRIOR FILING DATE: 2002-07-16

```
;  PRIOR APPLICATION NUMBER: US/10/264,237
;  CURRENT FILING DATE: 2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/16450
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,515
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 2876
;  SOFTWARE: PatentIn Ver. 3.1
;  SEQ ID NO 488
;  LENGTH: 562
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (34)..(34)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (473)..(473)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (480)..(480)
;  OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-488

Alignment Scores:
Pred. No.:               98.4               Length:        562
Score:                   9.00               Matches:       9
Percent Similarity:      100.0%             Conservative:  0
Best Local Similarity:   100.0%             Mismatches:    0
Query Match:             4.2%               Indels:        0
DB:                                         Gaps:          0

US-09-989-890-238 (1-212) x US-10-264-237-488 (1-562)

Qy      24 LeuProSerLeuAlaLeuLeuSer 32
           ||||||||||||||||||||||||
Db     355 TTGCCATCCTCTCTGGCCTTGTGTCC 381

RESULT 33
US-10-437-963-84925/c
;  Sequence 84925, Application US/10437963
;  Publication No. US20040123343A1
;  GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT: Kovalic, David K.
;  APPLICANT: Zhou, Yihua
;  APPLICANT: Cao, Yongwei
;  APPLICANT: Wu, Wei
;  APPLICANT: Boukharov, Andrey A.
;  APPLICANT: Barbazuk, Brad
;  APPLICANT: Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE: 2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
;  SEQ ID NO 84925
;  LENGTH: 568
;  TYPE: DNA
;  ORGANISM: Oryza sativa
;  FEATURE:
;  OTHER INFORMATION: Clone ID: PAT_MRT4530_84113C.1
US-10-437-963-84925

Alignment Scores:
Pred. No.:               99.3               Length:        568
Score:                   9.00               Matches:       9
Percent Similarity:      100.0%             Conservative:  0
Best Local Similarity:   100.0%             Mismatches:    0
Query Match:             4.2%               Indels:        0
```

```
;  PRIOR APPLICATION NUMBER: US/09/281,976
;  PRIOR FILING DATE: 1999-03-31
;  PRIOR APPLICATION NUMBER: 60/060,837
;  PRIOR FILING DATE: 1997-10-02
;  PRIOR APPLICATION NUMBER: 60/060,862
;  PRIOR FILING DATE: 1997-10-02
;  NUMBER OF SEQ ID NOS: 390
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 73
;  LENGTH: 546
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-979-183-73

Alignment Scores:
Pred. No.:               96.1               Length:        546
Score:                   9.00               Matches:       9
Percent Similarity:      100.0%             Conservative:  0
Best Local Similarity:   100.0%             Mismatches:    0
Query Match:             4.2%               Indels:        0
DB:                                         Gaps:          0

US-09-989-890-238 (1-212) x US-10-979-183-73 (1-546)

Qy     127 ThrProMetLeuSerSerLysAlaSer 135
           |||||||||||||||||||||||||||
Db     276 ACACCTATGCTCCTCCCTCCAAAGCTTCA 250

RESULT 31
US-09-764-875-478
;  Sequence 478, Application US/09764875
;  Publication No. US20040018969A1
;  GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PUZ02
;  CURRENT APPLICATION NUMBER: US/09/764,875
;  CURRENT FILING DATE: 2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1249
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 478
;  LENGTH: 549
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (34)
;  OTHER INFORMATION: n equals a,t,g, or c
US-09-764-875-478

Alignment Scores:
Pred. No.:               96.5               Length:        549
Score:                   9.00               Matches:       9
Percent Similarity:      100.0%             Conservative:  0
Best Local Similarity:   100.0%             Mismatches:    0
Query Match:             4.2%               Indels:        0
DB:                                         Gaps:          0

US-09-989-890-238 (1-212) x US-09-764-875-478 (1-549)

Qy      24 LeuProSerLeuAlaLeuLeuSer 32
           ||||||||||||||||||||||||
Db     349 TTGCCATCCTCTCTGGCCTTGTGTCC 375

RESULT 32
US-10-264-237-488
;  Sequence 488, Application US/10264237
;  Publication No. US20040009491A1
;  GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA131P1
```

```
DB:                          7             Gaps:        0

US-09-989-890-238 (1-212) x US-10-437-963-84925 (1-568)

Qy         33 ArgProSerProProAlaAla 41
              |||||||||||||||||||||
Db        412 CGGCCTCTCTCCCCACCGCCGGCCGCT 386

RESULT 34
US-09-925-065A-468689/c
; Sequence 468689, Application US/09925065A
; Publication No. US2005022817ZA9
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 468689
;   LENGTH: 637
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-468689

Alignment Scores:
Pred. No.:                    109          Length:        637
Score:                       9.00          Matches:       9
Percent Similarity:        100.0%          Conservative:  0
Best Local Similarity:     100.0%          Mismatches:    0
Query Match:                 4.2%          Indels:        0
DB:                            4           Gaps:          0

US-09-989-890-238 (1-212) x US-09-925-065A-468689 (1-637)

Qy         25 ProSerSerAlaAlaLeuLeuSerArg 33
              |||||||||||||||||||||
Db        497 CCGAGCTCCCTGGCTCTACTATCCAGG 471

RESULT 35
US-10-437-963-14833/c
; Sequence 14833, Application US/10437963
; Publication No. US2004012334A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
;   SEQ ID NO 14833
;   TYPE: DNA
```

```
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_20732C.1
US-10-437-963-14833

Alignment Scores:
Pred. No.:                    110          Length:        645
Score:                       9.00          Matches:       9
Percent Similarity:        100.0%          Conservative:  0
Best Local Similarity:     100.0%          Mismatches:    0
Query Match:                 4.2%          Indels:        0
DB:                            7           Gaps:          0

US-09-989-890-238 (1-212) x US-10-437-963-14833 (1-645)

Qy         33 ArgProLeuSerProProAlaAla 41
              ||||||||||||||||||||||||
Db         88 CGTCCACTTTCTCCGCCACCGGCAGCC 62

RESULT 36
US-09-925-065A-726141/c
; Sequence 726141, Application US/09925065A
; Publication No. US2005022817ZA9
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 726141
;   LENGTH: 652
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-726141

Alignment Scores:
Pred. No.:                    111          Length:        652
Score:                       9.00          Matches:       9
Percent Similarity:        100.0%          Conservative:  0
Best Local Similarity:     100.0%          Mismatches:    0
Query Match:                 4.2%          Indels:        4
DB:                                        Gaps:          0

US-09-989-890-238 (1-212) x US-09-925-065A-726141 (1-652)

Qy         28 LeuAlaLeuLeuSerArgProLeuSer 36
              ||||||||||||||||||||||
Db        576 TTAGCCCTTCTCTCTCGGCCTTTGTCT 550

RESULT 37
US-10-260-238-4042/c
; Sequence 4042, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
;   APPLICANT: Budworth, Paul R.
;   APPLICANT: Moughamer, Todd G.
;   APPLICANT: Briggs, Steven P.
;   APPLICANT: Cooper, Bret
;   APPLICANT: Glazebrook, Jane
;   APPLICANT: Goff, Stephen A.
```

```
;   APPLICANT: Katagiri, Fumiyaki
;   APPLICANT: Kreps, Joel
;   APPLICANT: Provart, Nicholas
;   APPLICANT: Ricke, Darrell
;   APPLICANT: Zhu, Tong
;   TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
;   FILE REFERENCE: 60111-NP
;   CURRENT APPLICATION NUMBER: US/10/260,238
;   CURRENT FILING DATE: 2002-09-26
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,448
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,277
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/370,620
;   PRIOR FILING DATE: 2002-04-04
;   NUMBER OF SEQ ID NOS: 6077
;   SEQ ID NO 4042
;     LENGTH: 717
;     TYPE: DNA
;     ORGANISM: Triticum aestivum
US-10-260-238-4042

Alignment Scores:
Pred. No.:              121             Length:          717
Score:                  9.00            Matches:         9
Percent Similarity:     100.0%          Conservative:    0
Best Local Similarity:  100.0%          Mismatches:      0
Query Match:            4.2%            Indels:          0
DB:                                     Gaps:            0

US-09-989-890-238 (1-212) x US-10-260-238-4042 (1-717)

Qy       24 LeuProSerLeuAlaLeuLeuSer 32
            ||||||||||||||||||||||||
Db      366 CTGCCTTCTTCATTGGCGCTGTTCG 340

RESULT 38
US-10-260-238-4040/c
; Sequence 4040, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
;   APPLICANT: Budworth, Paul R.
;   APPLICANT: Moughamer, Todd G.
;   APPLICANT: Briggs, Steven P.
;   APPLICANT: Cooper, Bret
;   APPLICANT: Glazebrook, Jane
;   APPLICANT: Goff, Stephen A.
;   APPLICANT: Katagiri, Fumiyaki
;   APPLICANT: Kreps, Joel
;   APPLICANT: Provart, Nicholas
;   APPLICANT: Ricke, Darrell
;   APPLICANT: Zhu, Tong
;   TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
;   FILE REFERENCE: 60111-NP
;   CURRENT APPLICATION NUMBER: US/10/260,238
;   CURRENT FILING DATE: 2002-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,448
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,277
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/370,620
;   PRIOR FILING DATE: 2002-04-04
;   NUMBER OF SEQ ID NOS: 6077
;   SEQ ID NO 4040
;     LENGTH: 737
;     TYPE: DNA
;     ORGANISM: Triticum aestivum
US-10-260-238-4040

Alignment Scores:
Pred. No.:              124             Length:          737
Score:                  9.00            Matches:         9
Percent Similarity:     100.0%          Conservative:    0
```

```
Best Local Similarity:  100.0%          Mismatches:      0
Query Match:            4.2%            Indels:          0
DB:                                     Gaps:            0

US-09-989-890-238 (1-212) x US-10-260-238-4040 (1-737)

Qy       24 LeuProSerSerLeuAlaLeuLeuSer 32
            |||||||||||||||||||||||||
Db      427 CTGCCTTCTTCGTTGGCGCTGTTGTCG 401

RESULT 39
US-10-296-115-280
; Sequence 280, Application US/1029615
; Publication No. US20040053248A1
; GENERAL INFORMATION:
;   APPLICANT: Hyseq Inc
;   TITLE OF INVENTION: No. US20040053248A1e1 Nucleic Acids and Polypeptides
;   FILE REFERENCE: 784PCT
;   CURRENT APPLICATION NUMBER: US/10/296,115
;   CURRENT FILING DATE: 2002-11-18
;   PRIOR APPLICATION NUMBER: US09/488,725
;   PRIOR FILING DATE: 2000-01-21
;   PRIOR APPLICATION NUMBER: US09/552,317
;   PRIOR FILING DATE: 2000-04-25
;   NUMBER OF SEQ ID NOS: 1478
;   SEQ ID NO 280
;     LENGTH: 792
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-10-296-115-280

Alignment Scores:
Pred. No.:              131             Length:          792
Score:                  9.00            Matches:         9
Percent Similarity:     100.0%          Conservative:    0
Best Local Similarity:  100.0%          Mismatches:      0
Query Match:            4.2%            Indels:          0
DB:                                     Gaps:            0

US-09-989-890-238 (1-212) x US-10-296-115-280 (1-792)

Qy       24 LeuProSerSerLeuAlaLeuLeuSer 32
            |||||||||||||||||||||||||
Db      755 TTGCCATCCTCTCGGCCTTGTTGTCC 781

RESULT 40
US-10-260-238-4041/c
; Sequence 4041, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
;   APPLICANT: Budworth, Paul R.
;   APPLICANT: Moughamer, Todd G.
;   APPLICANT: Briggs, Steven P.
;   APPLICANT: Cooper, Bret
;   APPLICANT: Glazebrook, Jane
;   APPLICANT: Goff, Stephen A.
;   APPLICANT: Katagiri, Fumiyaki
;   APPLICANT: Kreps, Joel
;   APPLICANT: Provart, Nicholas
;   APPLICANT: Ricke, Darrell
;   APPLICANT: Zhu, Tong
;   TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
;   FILE REFERENCE: 60111-NP
;   CURRENT APPLICATION NUMBER: US/10/260,238
;   CURRENT FILING DATE: 2002-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,448
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,277
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/370,620
;   PRIOR FILING DATE: 2002-04-04
;   NUMBER OF SEQ ID NOS: 6077
;   SEQ ID NO 4041
```

```
;   LENGTH: 824
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-10-260-238-4041
```

Alignment Scores:

Pred. No.:	136	Length:	824
Score:	9.00	Matches:	9
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	4.2%	Indels:	0
DB:		Gaps:	0

US-09-989-890-238 (1-212) x US-10-260-238-4041 (1-824)

```
Qy   24 LeuProSerSerLeuAlaAlaLeuLeuSer 32
            ||||||||||||||||||||||||
Db  455 CTGCCTTCTTCATTCGGCGGCTTGTTGTCG 429
```

RESULT 41
US-10-437-963-86193
```
; Sequence 86193, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
;   SEQ ID NO 86193
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_85257C.1
US-10-437-963-86193
```

Alignment Scores:

Pred. No.:	145	Length:	891
Score:	9.00	Matches:	9
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	4.2%	Indels:	0
DB:		Gaps:	0

US-09-989-890-238 (1-212) x US-10-437-963-86193 (1-891)

```
Qy   52 AlaGlyLeuProSerAlaSerAlaAla 60
            ||||||||||||||||||||
Db  459 GCCGGACTTCCTTCGGCTTCGGCCGCG 485
```

RESULT 42
US-10-437-963-34733/c
```
; Sequence 34733, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
;   SEQ ID NO 34733
;   LENGTH: 963
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_3871C.1
US-10-437-963-34733
```

Alignment Scores:

Pred. No.:	155	Length:	963
Score:	9.00	Matches:	9
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	4.2%	Indels:	0
DB:		Gaps:	0

US-09-989-890-238 (1-212) x US-10-437-963-34733 (1-963)

```
Qy   37 ProProProAlaAlaCysSerGlyAsp 45
            ||||||||||||||||||||
Db  487 CCTCCACCAGCAGCATGCTCTGGCGAC 461
```

RESULT 43
US-10-087-080-31/c
```
; Sequence 31, Application US/10087080
; Publication No. US20030235820A1
; GENERAL INFORMATION:
;   APPLICANT: Mack, David H.
;   APPLICANT: Markowitz, Sanford David
;   APPLICANT: Bos Biotechnology, Inc.
;   APPLICANT: Case Western Reserve University
;   TITLE OF INVENTION: No. US20030235820A1el Methods of Diagnosis of Metastatic Colorec
;   TITLE OF INVENTION: Cancer, Compositions and Methods of Screening for
;   TITLE OF INVENTION: Modulators of Metastatic Colorectal Cancer
;   FILE REFERENCE: 018501-000840US
;   CURRENT APPLICATION NUMBER: US/10/087,080
;   CURRENT FILING DATE: 2002-10-25
;   PRIOR APPLICATION NUMBER: US 60/272,206
;   PRIOR FILING DATE: 2001-02-27
;   PRIOR APPLICATION NUMBER: US 60/281,149
;   PRIOR FILING DATE: 2001-04-02
;   PRIOR APPLICATION NUMBER: US 60/284,555
;   PRIOR FILING DATE: 2001-04-17
;   NUMBER OF SEQ ID NOS: 41
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 31
;   LENGTH: 1209
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: winged helix/forkhead transcription factor (HFH1)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (1161)
;   OTHER INFORMATION: n = g, a, c or t
US-10-087-080-31
```

Alignment Scores:

Pred. No.:	187	Length:	1209
Score:	9.00	Matches:	9
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	4.2%	Indels:	0
DB:		Gaps:	6

US-09-989-890-238 (1-212) x US-10-087-080-31 (1-1209)

```
Qy      59 AlaAlaAlaGlyIleAlaSerSerAla 67
           ||||||||||||||||||||||||||
Db     242 GCTGCTGCCGGGATCGCCTCCTCCGCG 216
```

RESULT 46
US-10-650-112-9/c
; Sequence 9, Application US/10650112
; Publication No. US20040110712A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR TREATING PATIENTS AND IDENTIFYING THERAPEUTICS
; FILE REFERENCE: CWRU-P01-044
; CURRENT APPLICATION NUMBER: US/10/650,112
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 10/274,177
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: 10/229,245
; PRIOR FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/406,296
; PRIOR FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1161)..(1161)
; OTHER INFORMATION: n is a,t,g or c
US-10-650-112-9

Alignment Scores:
Pred. No.:   187   Length: 1209
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 7   Gaps: 0

US-09-989-890-238 (1-212) x US-10-650-112-9 (1-1209)

```
Qy      59 AlaAlaAlaGlyIleAlaSerSerAla 67
           ||||||||||||||||||||||||||
Db     242 GCTGCTGCCGGGATCGCCTCCTCCGCG 216
```

RESULT 47
US-11-041-788-9/c
; Sequence 9, Application US/11041788
; Publication No. US20050233353A1
; GENERAL INFORMATION:
; APPLICANT: Markowitz, Sanford D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P04-003
; CURRENT APPLICATION NUMBER: US/11/041,788
; CURRENT FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: US 10/649,591
; PRIOR FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: US 10/274,177
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US 10/229,345
; PRIOR FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: US 10/087,080
; PRIOR FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 60/284,555
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 60/281,149
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: US 60/272,206
; PRIOR FILING DATE: 2001-02-27
; NUMBER OF SEQ ID NOS: 22
```
Qy      59 AlaAlaAlaGlyIleAlaSerSerAla 67
           ||||||||||||||||||||||||||
Db     242 GCTGCTGCCGGGATCGCCTCCTCCGCG 216
```

RESULT 44
US-10-229-345-9/c
; Sequence 9, Application US/10229345
; Publication No. US20040038220A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/229,345
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1161)..(1161)
; OTHER INFORMATION: n=a, c, g, or t
US-10-229-345-9

Alignment Scores:
Pred. No.:   187   Length: 1209
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 7   Gaps: 0

US-09-989-890-238 (1-212) x US-10-229-345-9 (1-1209)

```
Qy      59 AlaAlaAlaGlyIleAlaSerSerAla 67
           ||||||||||||||||||||||||||
Db     242 GCTGCTGCCGGGATCGCCTCCTCCGCG 216
```

RESULT 45
US-10-274-177-9/c
; Sequence 9, Application US/10274177
; Publication No. US20040038225A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/274,177
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US/10/229,345
; PRIOR FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1161)..(1161)
; OTHER INFORMATION: n=a, c, g, or t
US-10-274-177-9

Alignment Scores:
Pred. No.:   187   Length: 1209
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 7   Gaps: 0

US-09-989-890-238 (1-212) x US-10-274-177-9 (1-1209)

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1161
; OTHER INFORMATION: n = A,T,C or G
US-11-041-788-9
```

Alignment Scores:
Pred. No.:   187   Length: 1209
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-11-041-788-9 (1-1209)

```
Qy    59 AlaAlaAlaGlyIleAlaSerSerAla 67
         |||||||||||||||||||||||||||
Db   242 GCTGCTGCCGGGGATCGCCTCCTCCGCG 216
```

RESULT 48
US-10-650-112-25/c
; Sequence 25, Application US/10650112
; Publication No. US20040110712A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR TREATING PATIENTS AND IDENTIFYING THERAPEUTICS
; FILE REFERENCE: CWRU-P01-044
; CURRENT APPLICATION NUMBER: US/10/650,112
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 10/274,177
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: 10/229,245
; PRIOR FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/406,296
; PRIOR FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: HUMAN FOXQ1
US-10-650-112-25

Alignment Scores:
Pred. No.:   187   Length: 1212
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-650-112-25 (1-1212)

```
Qy    59 AlaAlaAlaGlyIleAlaSerSerAla 67
         |||||||||||||||||||||||||||
Db   245 GCTGCTGCCGGGGATCGCCTCCTCCGCG 219
```

RESULT 49
US-10-756-149-15/c
; Sequence 15, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file

```
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1448
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1393)..(1393)
; OTHER INFORMATION: n is a, c, g, or t
US-10-756-149-15
```

Alignment Scores:
Pred. No.:   218   Length: 1448
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-756-149-15 (1-1448)

```
Qy    59 AlaAlaAlaGlyIleAlaSerSerAla 67
         |||||||||||||||||||||||||||
Db   474 GCTGCTGCCGGGGATCGCCTCCTCCGCG 448
```

RESULT 50
US-09-764-875-179
; Sequence 179, Application US/0976 4875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ02
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 179
; LENGTH: 1564
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-875-179

Alignment Scores:
Pred. No.:   232   Length: 1564
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 3
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-764-875-179 (1-1564)

```
Qy    24 LeuProSerSerLeuAlaLeuLeuSer 32
         ||||||||||||||||||||||||||
Db   156 TTGCCATCCCTCTGGCCTTGTTGTCC 182
```

RESULT 51
US-10-471-010-1
; Sequence 1, Application US/10471010
; Publication No. US2004018527A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001044-US
; CURRENT APPLICATION NUMBER: US/10/471,010
; CURRENT FILING DATE: 2003-09-08

```
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 1
;  LENGTH: 1644
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-471-010-1

Alignment Scores:
Pred. No.:                                   Length:         1644
Score:                      242              Matches:           9
Percent Similarity:         100.0%           Conservative:      0
Best Local Similarity:      100.0%           Mismatches:        0
Query Match:                4.2%             Indels:            0
DB:                         8                Gaps:              0

US-09-989-890-238 (1-212) x US-10-471-010-1 (1-1644)

Qy       24 LeuProSerLeuAlaLeuLeuSer 32
            ||||||||||||||||||||||||
Db     1511 TTGCCATCCTCTCTGGCCTTGTTGTCC 1537


RESULT 52
US-10-332-447-61
; Sequence 61, Application US/10332447
; Publication No. US20040053258A1
; GENERAL INFORMATION:
;   APPLICANT: INCYTE GENOMICS, INC.; RAUMANN, Brigitte E.;
;   APPLICANT: THORNTON, Michael; DING, Li; YUE, Henry;
;   APPLICANT: TANG, Y.Tom; HARLAND, Lee; BURFORD, Neil;
;   APPLICANT: GREENE, Barrie D.; SANJANWALA, Madhu S.;
;   APPLICANT: BAUGHN, Mariah R.; YAO, Monique G.; YANG, Junming;
;   APPLICANT: ARVIZU, Chandra S.; GANDHI, Ameena R.;
;   APPLICANT: HAFALIA, April J.A.; TRIBOULEY, Catherine M.;
;   APPLICANT: WALIA, Narinder K.; AU-YOUNG, Janice;
;   APPLICANT: WALSH, Roderick T.; RAMKUMAR, Jayalaxmi;
;   APPLICANT: LU, Yan; LU, Dyung Aina M.; AZIMZAI, Yalda;
;   APPLICANT: LAL, Preeti; ELLIOTT, Vicki S.; NGUYEN, Danniel B.;
;   APPLICANT: XU, Yuming; SEILHAMER, Jeffrey J.; BOROWSKY, Mark L.;
;   APPLICANT: KHAN, Farrah A.; KEARNEY, Liam; THANGAVELU, Kavitha;
;   APPLICANT: DAS, Debopriya; POLICKY, Jennifer L.
;   TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
;   FILE REFERENCE: PI-0149 USN
;   CURRENT APPLICATION NUMBER: US/10/332,447
;   CURRENT FILING DATE:  2003-01-07
;   PRIOR APPLICATION NUMBER: US 60/216,547
;   PRIOR FILING DATE: 2000-07-07
;   PRIOR APPLICATION NUMBER: US 60/218,232
;   PRIOR FILING DATE: 2000-07-14
;   PRIOR APPLICATION NUMBER: US 60/220,112
;   PRIOR FILING DATE: 2000-07-21
;   PRIOR APPLICATION NUMBER: US 60/221,839
;   PRIOR FILING DATE: 2000-07-28
;   NUMBER OF SEQ ID NOS: 64
;   SOFTWARE: PERL Program
;   SEQ ID NO 1818
;   LENGTH: 1818
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US2004005325 8A1 168827CB1
US-10-332-447-61

Alignment Scores:
Pred. No.:                                   Length:         1818
Score:                      263              Matches:           9
Percent Similarity:         100.0%           Conservative:      0
Best Local Similarity:      100.0%           Mismatches:        0
Query Match:                4.2%             Indels:            0
DB:                         7                Gaps:              0

US-09-989-890-238 (1-212) x US-10-332-447-61 (1-1818)

Qy       24 LeuProSerLeuAlaLeuLeuSer 32
            ||||||||||||||||||||||||
Db     1685 TTGCCATCCTCTCTGGCCTTGTTGTCC 1711


RESULT 53
US-10-437-963-90203
; Sequence 90203, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE:  2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
;   SEQ ID NO 90203
;   LENGTH: 1923
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_88895C.1
US-10-437-963-90203

Alignment Scores:
Pred. No.:                                   Length:          1923
Score:                      276              Matches:            9
Percent Similarity:         100.0%           Conservative:       0
Best Local Similarity:      100.0%           Mismatches:         0
Query Match:                4.2%             Indels:             0
DB:                         7                Gaps:               0

US-09-989-890-238 (1-212) x US-10-437-963-90203 (1-1923)

Qy       32 SerArgProLeuSerProProProAla 40
            |||||||||||||||||||||||||||
Db      522 AGCCGGCCTCTATCTCCGGCCGCCCGCG 548


RESULT 54
US-10-369-493-26078/c
; Sequence 26078, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Hinkle, Gregory J.
;   APPLICANT: Slater, Steven C.
;   APPLICANT: Goldman, Barry S.
;   APPLICANT: Chen, Xianfeng
;   TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;   TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
;   FILE REFERENCE: 38-10(52052)B
;   CURRENT APPLICATION NUMBER: US/10/369,493
;   CURRENT FILING DATE:  2003-02-28
;   PRIOR APPLICATION NUMBER: US 60/360,039
;   PRIOR FILING DATE: 2002-02-21
;   NUMBER OF SEQ ID NOS: 47374
;   SEQ ID NO 26078
;   LENGTH: 2071
;   TYPE: DNA
;   ORGANISM: Schizosaccharomyces pombe
US-10-369-493-26078

Alignment Scores:
Pred. No.:                                   Length:          2071
Score:                      294              Matches:            9

US-09-989-890-238 (1-212) x US-10-332-447-61 (1-1818)
```

Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-369-493-26078 (1-2071)

Qy   175 ValLeuValAlaProArgSerThrIle 183
            |||||||||||||||||||||||||||
Db   1572 GTCTTAGTAGCCCCTGAAGCACGATT 1546

RESULT 55
US-10-332-447-62
; Sequence 62, Application US/10332447
; Publication No. US2004005325A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; RAUMANN, Brigitte E.;
; APPLICANT: THORNTON, Michael; DING, Li; YUE, Henry;
; APPLICANT: TANG, Y.Tom; HARLAND, Lee; BURFORD, Neil;
; APPLICANT: GREENE, Barrie D.; SANJANWALA, Madhu S.;
; APPLICANT: BAUGHN, Mariah R.; YAO, Monique G.; YANG, Junming;
; APPLICANT: ARVIZU, Chandra S.; GANDHI, Ameena R.;
; APPLICANT: HAFALIA, April J.A.; TRIBOULEY, Catherine M.;
; APPLICANT: WALIA, Narinder K.; AU-YOUNG, Janice;
; APPLICANT: WALSH, Roderick T.; RAMKUMAR, Jayalaxmi;
; APPLICANT: LU, Yan; LU, Dyung Aina M.; AZIMZAI, Yalda;
; APPLICANT: LAL, Preeti; ELLIOTT, Vicki S.; NGUYEN, Danniel B.;
; APPLICANT: XU, Yuning; SEILHAMER, Jeffrey J.; BOROWSKY, Mark L.;
; APPLICANT: KHAN, Farrah A.; KEARNEY, Liam; THANGAVELU, Kavitha;
; APPLICANT: DAS, Debopriya; POLICKY, Jennifer L.
; TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
; FILE REFERENCE: PI-0149 USN
; CURRENT APPLICATION NUMBER: US/10/332,447
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/216,547
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/218,232
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/220,112
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/221,839
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PERL Program
; SEQ ID NO 62
; LENGTH: 2245
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US2004005325A1 7472734CB1
US-10-332-447-62

Alignment Scores:
Pred. No.:   314   Length: 2245
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-332-447-62 (1-2245)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
            |||||||||||||||||||||||||||
Db   2112 TTGGCATCCTCTCTCGGCCCTGTTGTCC 2138

RESULT 56
US-10-450-763-15603/c
; Sequence 15603, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 15603
; LENGTH: 2256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (2139)..(1663)
; OTHER INFORMATION: 100% homologous to Escherichia coli Spermidine/putrescine
; OTHER INFORMATION: transport protein A,accession number D90748,Smith-Waterman Score
; OTHER INFORMATION: 834.
US-10-450-763-15603

Alignment Scores:
Pred. No.:   315   Length: 2256
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-450-763-15603 (1-2256)

Qy   54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
            ||||||||||||||||||||||||||||||
Db   337 CTTCCTTCAGCGTCAGCAGCAGCTGGA 311

RESULT 57
US-10-437-963-4805
; Sequence 4805, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 4805
; LENGTH: 2403
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_11655C.1
US-10-437-963-4805

Alignment Scores:
Pred. No.:   332   Length: 2403
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 7

US-09-989-890-238 (1-212) x US-10-437-963-4805 (1-2403)

Qy   25 ProSerLeuAlaLeuLeuSerArg 33
      ||||||||||||||||||||
Db   80 CCTTCCTCCTTGCATTGCTCTCTCGG 106

RESULT 58
US-09-925-065A-691773
; Sequence 691773, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR APPLICATION FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR APPLICATION FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR APPLICATION FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR APPLICATION FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR APPLICATION FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 691773
; LENGTH: 2520
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1550_1551
; OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-691773

Alignment Scores:
Pred. No.:   346   Length: 2520
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
Db: 4   Gaps: 0
US-09-925-065A-691773

US-09-989-890-238 (1-212) x US-09-925-065A-691773 (1-2520)

Qy   23 AlaLeuProSerSerLeuAlaLeuLeu 31
      ||||||||||||||||||||||
Db   93 GCTCTTCCAAGTAGTCTTGCTCTGCTG 119

RESULT 59
US-10-437-963-86961
; Sequence 86961, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86961

; LENGTH: 3795
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85954C.1
US-10-437-963-86961

Alignment Scores:
Pred. No.:   488   Length: 3795
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
Db: 7   Gaps: 0

US-09-989-890-238 (1-212) x US-10-437-963-86961 (1-3795)

Qy   37 ProProAlaAlaCysSerGlyAsp 45
      |||||||||||||||||||
Db 2396 CCTCCACCAGCAGCATGCTCTGGCGAC 2422

RESULT 60
US-11-097-143-23948
; Sequence 23948, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR APPLICATION FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR APPLICATION FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR APPLICATION FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR APPLICATION FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR APPLICATION FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR APPLICATION FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR APPLICATION FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR APPLICATION FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23948
; LENGTH: 3840
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-23948

Alignment Scores:
Pred. No.:   492   Length: 3840
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
Db: 10   Gaps: 0

US-09-989-890-238 (1-212) x US-11-097-143-23948 (1-3840)

Qy   54 LeuProSerAlaSerAlaAlaAlaGly 62
      |||||||||||||||||||||
Db 567 CTACCATCCCCCAGCGCCCCGCCCCGGT 593

RESULT 61

US-11-097-143-23947/c
; Sequence 23947, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23947
;   LENGTH: 5906
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-23947

Alignment Scores:
Pred. No.:   706   Length: 5906
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-11-097-143-23947 (1-5906)

Qy   54 LeuProSerAlaSerAlaAlaAlaAlaGly 62
             ||||||||||||||||||||||||||||||
Db   4274 CTACCATCCGCCAGCGCCGCCGCCGGT 4248

RESULT 62
US-09-764-860-960/c
; Sequence 960, Application US/09764860
; Patent No. US2002009495A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 960
;   LENGTH: 7880
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-860-960

Alignment Scores:
Pred. No.:   899   Length: 7880
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0

Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-764-860-960 (1-7880)

Qy   25 ProSerLeuAlaLeuSerArg 33
             |||||||||||||||||||||
Db   4716 CCGAGCTCCCTGGCTCTACTATCCAGG 4690

RESULT 63
US-09-764-860-961/c
; Sequence 961, Application US/09764860
; Patent No. US2002009495A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 961
;   LENGTH: 7880
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-860-961

Alignment Scores:
Pred. No.:   899   Length: 7880
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-764-860-961 (1-7880)

Qy   25 ProSerLeuAlaLeuSerArg 33
             |||||||||||||||||||||
Db   4716 CCGAGCTCCCTGGCTCTACTATCCAGG 4690

RESULT 64
US-10-074-095-960/c
; Sequence 960, Application US/10074095
; Publication No. US2003007704A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C1
; CURRENT APPLICATION NUMBER: US/10/074,095
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 09/764,860
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: 60/225,757
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/226,868
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/216,647
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: 60/225,267
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/216,880
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: 60/225,270
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/251,869
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/235,834
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/234,274
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: 60/234,223
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: 60/228,924
PRIOR FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/224,518
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/224,519
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,964
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/249,299
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/236,327
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: 60/229,343
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/236,802

PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,037
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,040
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/240,960
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/239,935
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/239,937
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/241,787
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,759
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/227,182
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,214
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,836
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/230,438
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/215,135
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 60/225,266
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/249,218
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,208
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,213
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,212
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,207
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,245
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,244
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,217
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,211
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,215
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,264
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,214
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,297
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/232,400
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/231,242
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,081
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,080
PRIOR FILING DATE: 2000-09-08

```
,,  PRIOR APPLICATION NUMBER: 60/226,868
,,  PRIOR FILING DATE: 2000-08-22
,,  PRIOR APPLICATION NUMBER: 60/216,647
,,  PRIOR FILING DATE: 2000-07-07
,,  PRIOR APPLICATION NUMBER: 60/225,267
,,  PRIOR FILING DATE: 2000-08-14
,,  PRIOR APPLICATION NUMBER: 60/216,880
,,  PRIOR FILING DATE: 2000-07-07
,,  PRIOR APPLICATION NUMBER: 60/225,270
,,  PRIOR FILING DATE: 2000-08-14
,,  PRIOR APPLICATION NUMBER: 60/251,869
,,  PRIOR FILING DATE: 2000-12-08
,,  PRIOR APPLICATION NUMBER: 60/235,834
,,  PRIOR FILING DATE: 2000-09-27
,,  PRIOR APPLICATION NUMBER: 60/234,274
,,  PRIOR FILING DATE: 2000-09-21
,,  PRIOR APPLICATION NUMBER: 60/234,223
,,  PRIOR FILING DATE: 2000-09-21
,,  PRIOR APPLICATION NUMBER: 60/228,924
,,  PRIOR FILING DATE: 2000-08-30
,,  PRIOR APPLICATION NUMBER: 60/224,518
,,  PRIOR FILING DATE: 2000-08-14
,,  PRIOR APPLICATION NUMBER: 60/236,369
,,  PRIOR FILING DATE: 2000-09-29
,,  PRIOR APPLICATION NUMBER: 60/224,519
,,  PRIOR FILING DATE: 2000-08-14
,,  PRIOR APPLICATION NUMBER: 60/220,964
,,  PRIOR FILING DATE: 2000-07-26
,,  PRIOR APPLICATION NUMBER: 60/241,809
,,  PRIOR FILING DATE: 2000-10-20
,,  PRIOR APPLICATION NUMBER: 60/249,299
,,  PRIOR FILING DATE: 2000-11-17
,,  PRIOR APPLICATION NUMBER: 60/236,327
,,  PRIOR FILING DATE: 2000-09-29
,,  PRIOR APPLICATION NUMBER: 60/241,785
,,  PRIOR FILING DATE: 2000-10-20
,,  PRIOR APPLICATION NUMBER: 60/244,617
,,  PRIOR FILING DATE: 2000-11-01
,,  PRIOR APPLICATION NUMBER: 60/225,268
,,  PRIOR FILING DATE: 2000-08-14
,,  PRIOR APPLICATION NUMBER: 60/236,368
,,  PRIOR FILING DATE: 2000-09-29
,,  PRIOR APPLICATION NUMBER: 60/251,856
,,  PRIOR FILING DATE: 2000-12-08
,,  PRIOR APPLICATION NUMBER: 60/251,868
,,  PRIOR FILING DATE: 2000-12-08
,,  PRIOR APPLICATION NUMBER: 60/229,344
,,  PRIOR FILING DATE: 2000-09-01
,,  PRIOR APPLICATION NUMBER: 60/234,997
,,  PRIOR FILING DATE: 2000-09-25
,,  PRIOR APPLICATION NUMBER: 60/229,343
,,  PRIOR FILING DATE: 2000-09-01
,,  PRIOR APPLICATION NUMBER: 60/229,513
,,  PRIOR FILING DATE: 2000-09-05
,,  PRIOR APPLICATION NUMBER: 60/231,413
,,  PRIOR FILING DATE: 2000-09-08
,,  PRIOR APPLICATION NUMBER: 60/229,509
,,  PRIOR FILING DATE: 2000-09-05
,,  PRIOR APPLICATION NUMBER: 60/236,367
,,  PRIOR FILING DATE: 2000-09-29
,,  PRIOR APPLICATION NUMBER: 60/237,039
,,  PRIOR FILING DATE: 2000-10-02
,,  PRIOR APPLICATION NUMBER: 60/237,038
,,  PRIOR FILING DATE: 2000-10-02
,,  PRIOR APPLICATION NUMBER: 60/236,370
,,  PRIOR FILING DATE: 2000-09-29
,,  PRIOR APPLICATION NUMBER: 60/236,802
,,  PRIOR FILING DATE: 2000-10-02
,,  PRIOR APPLICATION NUMBER: 60/237,037
,,  PRIOR FILING DATE: 2000-08-14
```

```
,,  PRIOR APPLICATION NUMBER: 60/231,414
,,  PRIOR FILING DATE: 2000-09-08
,,  PRIOR APPLICATION NUMBER: 60/231,244
,,  PRIOR FILING DATE: 2000-09-08
,,  PRIOR APPLICATION NUMBER: 60/233,064
,,  PRIOR FILING DATE: 2000-09-14
,,  PRIOR APPLICATION NUMBER: 60/233,063
,,  PRIOR FILING DATE: 2000-09-14
,,  PRIOR APPLICATION NUMBER: 60/232,397
,,  PRIOR FILING DATE: 2000-09-14
,,  PRIOR APPLICATION NUMBER: 60/232,399
,,  PRIOR FILING DATE: 2000-09-14
,,  PRIOR APPLICATION NUMBER: 60/232,401
,,  PRIOR FILING DATE: 2000-09-14
,,  PRIOR APPLICATION NUMBER: 60/241,808
,,  PRIOR FILING DATE: 2000-10-20
,,  PRIOR APPLICATION NUMBER: 60/241,826
,,  PRIOR FILING DATE: 2000-10-20
,,  PRIOR APPLICATION NUMBER: 60/241,786
,,  PRIOR FILING DATE: 2000-10-20
,,  PRIOR APPLICATION NUMBER: 60/241,221
,,  PRIOR FILING DATE: 2000-10-20
,,  PRIOR APPLICATION NUMBER: 60/246,475
,,  PRIOR FILING DATE: 2000-11-08
,,  PRIOR APPLICATION NUMBER: 60/231,243
,,  PRIOR FILING DATE: 2000-09-08

Alignment Scores:
Pred. No.:                     899          Length:        7880
Score:                        9.00          Matches:          9
Percent Similarity:         100.0%          Conservative:     0
Best Local Similarity:      100.0%          Mismatches:       0
Query Match:                  4.2%          Indels:           0
DB:                              5          Gaps:             0

US-09-989-890-238 (1-212) x US-10-074-095-960 (1-7880)

Qy        25 ProSerSerLeuAlaLeuLeuSerArg  33
              |||||||||||||||||||||||||||
Db      4716 CCGAGCTCCCTGGCTCTACTATCCAGG 4690

RESULT 65
US-10-074-095-961/c
; Sequence 961, Application US/10074095
; Publication No. US20030077704A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC008C1
;  CURRENT APPLICATION NUMBER: US/10/074,095
;  CURRENT FILING DATE: 2002-02-14
,,  PRIOR APPLICATION NUMBER: 09/764,860
,,  PRIOR FILING DATE: 2001-01-17
,,  PRIOR APPLICATION NUMBER: 60/179,065
,,  PRIOR FILING DATE: 2000-01-31
,,  PRIOR APPLICATION NUMBER: 60/180,628
,,  PRIOR FILING DATE: 2000-02-04
,,  PRIOR APPLICATION NUMBER: 60/214,886
,,  PRIOR FILING DATE: 2000-06-28
,,  PRIOR APPLICATION NUMBER: 60/217,487
,,  PRIOR FILING DATE: 2000-07-11
,,  PRIOR APPLICATION NUMBER: 60/225,758
,,  PRIOR FILING DATE: 2000-08-14
,,  PRIOR APPLICATION NUMBER: 60/220,963
,,  PRIOR FILING DATE: 2000-07-26
,,  PRIOR APPLICATION NUMBER: 60/217,496
,,  PRIOR FILING DATE: 2000-07-11
,,  PRIOR APPLICATION NUMBER: 60/225,447
,,  PRIOR FILING DATE: 2000-08-14
,,  PRIOR APPLICATION NUMBER: 60/218,290
,,  PRIOR FILING DATE: 2000-07-14
,,  PRIOR APPLICATION NUMBER: 60/225,757
,,  PRIOR FILING DATE: 2000-08-14
```

```
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08

; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08


Alignment Scores:
Pred. No.:                899                     Length:        7880
Score:                    9.00                    Matches:          9
Percent Similarity:       100.0%                  Conservative:     0
Best Local Similarity:    100.0%                  Mismatches:       0
Query Match:              4.2%                     Indels:           0
DB:                       5                        Gaps:             0

US-09-989-890-238 (1-212) x US-US-10-074-095-961 (1-7880)

Qy     25 ProSerLeuAlaLeuLeuSerArg 33
          ||||||||||||||||||||||||
Db   4716 CCGAGCTCCCTGGCTCTACTATCCAGG 4690

RESULT 66
US-10-212-872-960/c
; Sequence 960, Application US/10212872
; Publication No. US20030215893A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C2
; CURRENT APPLICATION NUMBER: US/10/212,872
; CURRENT FILING DATE: 2002-08-07
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 960
; LENGTH: 7880
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-212-872-960

Alignment Scores:
Pred. No.:                899                     Length:        7880
Score:                    9.00                    Matches:          9
Percent Similarity:       100.0%                  Conservative:     0
Best Local Similarity:    100.0%                  Mismatches:       0
Query Match:              4.2%                     Indels:           0
DB:                       6                        Gaps:             0

US-09-989-890-238 (1-212) x US-10-212-872-960 (1-7880)

Qy     25 ProSerLeuAlaLeuLeuSerArg 33
          ||||||||||||||||||||||||
Db   4716 CCGAGCTCCCTGGCTCTACTATCCAGG 4690

RESULT 67
```

US-10-212-872-961/c
; Sequence 961, Application US/10212872
; Publication No. US20030215893A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C2
; CURRENT APPLICATION NUMBER: US/10/212,872
; CURRENT FILING DATE: 2002-08-07
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 961
; LENGTH: 7880
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-212-872-961

Alignment Scores:
Pred. No.:   899   Length: 7880
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-212-872-961 (1-7880)

Qy   25 ProSerSerLeuAlaLeuLeuSerArg 33
           |||||||||||||||||||||||||
Db   4716 CCGAGCTCCCTGGCTCCTACTATCCAGG 4690

RESULT 68
US-10-156-761-411/c
; Sequence 411, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 411
; LENGTH: 10056
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(10056)
US-10-156-761-411

Alignment Scores:
Pred. No.:   1.1e+03   Length: 10056
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-156-761-411 (1-10056)

Qy   52 AlaGlyLeuProSerAlaSerAlaAla 60

Db   6155 GCGGGTTTGCCGTCGGCGAGCGCGCCG 6129
             |||||||||||||||||||||||||||

RESULT 69
US-09-919-172-97/c
; Sequence 97, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 97
; LENGTH: 10432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 2700132CB1
US-09-919-172-97

Alignment Scores:
Pred. No.:   1.14e+03   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-919-172-97 (1-10432)

Qy   29 AlaLeuLeuSerArgProLeuSerPro 37
           |||||||||||||||||||||||||
Db   5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 70
US-09-919-039-20/c
; Sequence 20, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 20
; LENGTH: 10432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2700132CB1
US-09-919-039-20

Alignment Scores:
Pred. No.:   1.14e+03   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-919-039-20 (1-10432)

RESULT 71
US-10-084-817-349/c
; Sequence 349, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Plon
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: PA-0046 US
; CURRENT APPLICATION NUMBER: US/10/084,817
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/270,784
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 365
; SOFTWARE: PERL Program
; SEQ ID NO 349
; LENGTH: 10432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030119009A1 2700132CB1
US-10-084-817-349

Alignment Scores:
Pred. No.:   1.14e+03   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 5   Gaps: 0

US-09-989-890-238 (1-212) x US-10-084-817-349 (1-10432)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 72
US-10-133-013-262/c
; Sequence 262, Application US/10133013
; Publication No. US20030166903A1
; GENERAL INFORMATION:
; APPLICANT: Astromoff, Anna
; APPLICANT: Bandman, Olga
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES ASSOCIATED WITH VASCULAR DISEASE
; FILE REFERENCE: PA-0049 US
; CURRENT APPLICATION NUMBER: US/10/133,013
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/287,067
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 271
; SOFTWARE: PERL Program
; SEQ ID NO 262
; LENGTH: 10432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030166903A1 2700132CB1
US-10-133-013-262

Alignment Scores:
Pred. No.:   1.14e+03   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-133-013-262 (1-10432)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 73
US-10-752-986-97/c
; Sequence 97, Application US/10752986
; Publication No. US20040253609A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/10/752,986
; CURRENT FILING DATE: 2004-01-06
; PRIOR APPLICATION NUMBER: US/09/919,172
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 97
; LENGTH: 10432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2700132CB1
US-10-752-986-97

Alignment Scores:
Pred. No.:   1.14e+03   Length: 10432
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x US-10-752-986-97 (1-10432)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 5063 GCTCTTCTTTCACGCCCACTTTCCCCA 5037

RESULT 74
US-10-723-860-5608/c
; Sequence 5608, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5608
; LENGTH: 11466
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-5608

Alignment Scores:
Pred. No.:
Score: 1.23e+03
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 4.2%
DB:

Length: 11466
Matches: 9
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x US-10-723-860-5608 (1-11466)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 4052 GCTCTTCTTTCACGCCCACTTTCCCCA 4026

RESULT 75
US-10-719-993-402/c
; Sequence 402, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;   ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 402
; LENGTH: 11598
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-402

Alignment Scores:
Pred. No.:
Score: 1.24e+03
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 4.2%
DB:

Length: 11598
Matches: 9
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x US-10-719-993-402 (1-11598)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 4232 GCTCTTCTTTCACGCCCACTTTCCCCA 4206

RESULT 76
US-10-719-993-403/c
; Sequence 403, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;   ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 403
; LENGTH: 11602
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-403

Alignment Scores:
Pred. No.:
Score: 1.24e+03
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 4.2%

Length: 11602
Matches: 9
Conservative: 0
Mismatches: 0
Indels: 0

DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-719-993-403 (1-11602)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 4236 GCTCTTCTTTCACGCCCACTTTCCCCA 4210

RESULT 77
US-10-739-946-1/c
; Sequence 1, Application US/10739946
; Publication No. US20040242519A1
; GENERAL INFORMATION:
; APPLICANT: Faustus Forschungs Cie. Translational Cancer Research GmbH
; TITLE OF INVENTION: Antisense oligonucleotides for treating proliferating cells
; FILE REFERENCE: PA32470US-01938
; CURRENT APPLICATION NUMBER: US/10/739,946
; CURRENT FILING DATE: 2003-12-19
; PRIOR APPLICATION NUMBER: DE 198 22 954.2
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 1
; LENGTH: 12493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (197)..(9964)
US-10-739-946-1

Alignment Scores:
Pred. No.:   Length: 12493
Score: 1.32e+03   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-739-946-1 (1-12493)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 78
US-10-252-157-263/c
; Sequence 263, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 263
; LENGTH: 12494
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US2003019O640A1 412661.2
US-10-252-157-263

Alignment Scores:
Pred. No.:   Length: 12494
Score: 1.32e+03   Matches: 9
Percent Similarity: 100.0%   Conservative: 0

Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-252-157-263 (1-12494)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 79
US-10-115-831-18/c
; Sequence 18, Application US/10115831
; Publication No. US20030219743A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Peiyan
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219743A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 792CIP2ADIV
; CURRENT APPLICATION NUMBER: US/10/115,831
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/667,298
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 178
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 18
; LENGTH: 12505
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (197)..(9967)
US-10-115-831-18

Alignment Scores:
Pred. No.:   1.32e+03   Length: 12505
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-115-831-18 (1-12505)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 80
US-09-918-624B-48/c
; Sequence 48, Application US/0918624B
; Publication No. US20030113720A1
; GENERAL INFORMATION:
; APPLICANT: Schebye, Xiao Min
; APPLICANT: Sornabae, Thierry
; TITLE OF INVENTION: CDNAS EXPRESSED IN ADIPOCYTE DIFFERENTIATION
; FILE REFERENCE: PA-0033 US
; CURRENT FILING DATE: 2002-12-03
; PRIOR APPLICATION NUMBER: 60/222,470
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PERL Program
; SEQ ID NO 48
; LENGTH: 12515
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank ID No. US20030113720A1 g415818
US-09-918-624B-48

Alignment Scores:
Pred. No.:   1.32e+03   Length: 12515
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-918-624B-48 (1-12515)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 81
US-09-968-007A-989/c
; Sequence 989, Application US/0968007A
; Publication No. US20040115625A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Usining Signa
; TITLE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-71
; CURRENT APPLICATION NUMBER: US/09/968,007A
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/60/237,172
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,173
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,278
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,294
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,295
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,316
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 1001
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 989
; LENGTH: 12515
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-968-007A-989

Alignment Scores:
Pred. No.:   1.32e+03   Length: 12515
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-968-007A-989 (1-12515)

Qy   29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106

RESULT 82
US-10-388-360-313/c
; Sequence 313, Application US/10388360
; Publication No. US20030225528A1
; GENERAL INFORMATION: GENOMIC HEALTH
; APPLICANT: Baker, Joffre B.
; APPLICANT: Cronin, Maureen T.
; APPLICANT: Kiefer, Michael C.

; APPLICANT: Shak, Steve
; APPLICANT: Walker, Michael Graham
; TITLE OF INVENTION: GENE EXPRESSION PROFILING IN BIOPSIED TUMOR TISSUES
; FILE REFERENCE: 39740-0001US
; CURRENT APPLICATION NUMBER: US/10/388,360
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: US 60/412,049
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/364,890
; PRIOR FILING DATE: 2002-03-13
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 313
;   LENGTH: 12515
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-388-360-313

Alignment Scores:
Pred. No.:   1.32e+03   Length: 12515
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-388-360-313 (1-12515)

Qy   29 AlaLeuLeuSerArgProLeuSerPro 37
              |||||||||||||||||||||||||||
Db   5132 GTCTTCTTTCACGCCCACTTCCCCA 5106

RESULT 83
US-10-719-993-405/c
; Sequence 405, Application US/10719993
; Publication No. US2004026584A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 405
;   LENGTH: 12515
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-405

Alignment Scores:
Pred. No.:   1.32e+03   Length: 12515
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x US-10-719-993-405 (1-12515)

Qy   29 AlaLeuLeuSerArgProLeuSerPro 37
              |||||||||||||||||||||||||||
Db   5132 GTCTTCTTTCACGCCCACTTCCCCA 5106

RESULT 84
US-10-843-641A-7459/c
; Sequence 7459, Application US/10843641A
; Publication No. US2005006445A1
; GENERAL INFORMATION:
; APPLICANT: Avalon Pharmaceuticals, Inc.
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; TITLE OF INVENTION: Signature Gene Sets

; FILE REFERENCE: 689290-189
; CURRENT APPLICATION NUMBER: US/10/843,641A
; CURRENT FILING DATE: 2004-05-12
; PRIOR APPLICATION NUMBER: US/09/873,367
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,436
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,832
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/964,824
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/09/967,768
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/09/968,007
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,347
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,708
; PRIOR FILING DATE: 2001-10-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 8447
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7459
;   LENGTH: 12515
;   TYPE: DNA
;   ORGANISM: Homo sapiens
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(12515)
;   OTHER INFORMATION: n=a,t,g or c
US-10-843-641A-7459

Alignment Scores:
Pred. No.:   1.32e+03   Length: 12515
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 9   Gaps: 0

US-09-989-890-238 (1-212) x US-10-843-641A-7459 (1-12515)

Qy   29 AlaLeuLeuSerArgProLeuSerPro 37
              |||||||||||||||||||||||||||
Db   5132 GTCTTCTTTCACGCCCACTTCCCCA 5106

RESULT 85
US-10-928-465-58/c
; Sequence 58, Application US/10928465
; Publication No. US2005008999A1
; GENERAL INFORMATION:
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: JACQUEMIER, JOCELYNE
; APPLICANT: GINESTIER, CHRISTOPHE
; APPLICANT: DEBONO, STEPHANE
; APPLICANT: BORIE, NATHALIE
; TITLE OF INVENTION: IDENTIFICATION OF AN ERBB2 GENE EXPRESSION SIGNATURE IN
; TITLE OF INVENTION: BREAST CANCERS
; FILE REFERENCE: 1308-P-03
; CURRENT APPLICATION NUMBER: US/10/928,465
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/498,497
; PRIOR FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 58
;   LENGTH: 12515
;   TYPE: DNA

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA sequence
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(12515)
;   OTHER INFORMATION: antigen identified by monoclonal antibody
;   OTHER INFORMATION: ki-67 (MKI67) gene.
US-10-928-465-58

Alignment Scores:
Pred. No.:             1.32e+03          Length:             12515
Score:                 9.00              Matches:            9
Percent Similarity:    100.0%            Conservative:       0
Best Local Similarity: 100.0%            Mismatches:         0
Query Match:           4.2%              Indels:             0
DB:                                      Gaps:               0

US-09-989-890-238 (1-212) x US-10-928-465-58 (1-12515)

Qy         29 AlaLeuSerArgProLeuSerPro 37
              ||||||||||||||||||||||||
Db       5132 GCTCTTCTTTCACGCCCACTTTCCCCA 5106


RESULT 86
US-10-071-766-49/c
; Sequence 49, Application US2002019266
; Publication No. US2002019267AA1
; GENERAL INFORMATION:
;   APPLICANT: Huei-Mei Chen
;   TITLE OF INVENTION: GENES EXPRESSED IN SENESCENCE
;   FILE REFERENCE: PA-0043 US
;   CURRENT APPLICATION NUMBER: US/10/071,766
;   CURRENT FILING DATE:  2002-02-07
;   NUMBER OF SEQ ID NOS: 144
;   SOFTWARE: PERL Program
; SEQ ID NO 49
;   LENGTH: 12632
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US2002019267AA1 412661.2
US-10-071-766-49

Alignment Scores:
Pred. No.:             1.33e+03          Length:             12632
Score:                 9.00              Matches:            9
Percent Similarity:    100.0%            Conservative:       0
Best Local Similarity: 100.0%            Mismatches:         0
Query Match:           4.2%              Indels:             0
DB:                                      Gaps:               5

US-09-989-890-238 (1-212) x US-10-071-766-49 (1-12632)

Qy         29 AlaLeuSerArgProLeuSerPro 37
              ||||||||||||||||||||||||
Db       5270 GCTCTTCTTTCACGCCCACTTTCCCCA 5244


RESULT 87
US-10-719-993-404/c
; Sequence 404, Application US/10719993
; Publication No. US2004026584A1
; GENERAL INFORMATION:
;   APPLICANT: CARGILL, Michele et al.
;   TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;   TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001496
;   CURRENT APPLICATION NUMBER: US/10/719,993
;   CURRENT FILING DATE:  2003-11-24
;   NUMBER OF SEQ ID NOS: 55342
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 404
```

```
;   LENGTH: 12678
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-404

Alignment Scores:
Pred. No.:             1.34e+03          Length:             12678
Score:                 9.00              Matches:            9
Percent Similarity:    100.0%            Conservative:       0
Best Local Similarity: 100.0%            Mismatches:         0
Query Match:           4.2%              Indels:             0
DB:                                      Gaps:               0

US-09-989-890-238 (1-212) x US-10-719-993-404 (1-12678)

Qy         29 AlaLeuSerArgProLeuSerPro 37
              ||||||||||||||||||||||||
Db       5312 GCTCTTCTTTCACGCCCACTTTCCCCA 5286


RESULT 88
US-10-719-993-6886/c
; Sequence 6886, Application US/10719993
; Publication No. US2004026584A1
; GENERAL INFORMATION:
;   APPLICANT: CARGILL, Michele et al.
;   TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;   TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001496
;   CURRENT APPLICATION NUMBER: US/10/719,993
;   CURRENT FILING DATE:  2003-11-24
;   NUMBER OF SEQ ID NOS: 55342
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6886
;   LENGTH: 41736
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6886

Alignment Scores:
Pred. No.:             3.63e+03          Length:             41736
Score:                 9.00              Matches:            9
Percent Similarity:    100.0%            Conservative:       0
Best Local Similarity: 100.0%            Mismatches:         0
Query Match:           4.2%              Indels:             0
DB:                                      Gaps:               8

US-09-989-890-238 (1-212) x US-10-719-993-6886 (1-41736)

Qy         29 AlaLeuSerArgProLeuSerPro 37
              ||||||||||||||||||||||||
Db      25490 GCTCTTCTTTCACGCCCACTTTCCCCA 25464


RESULT 89
US-10-162-846-11
; Sequence 11, Application US/10162846
; Publication No. US2003022451A1
; GENERAL INFORMATION:
;   APPLICANT: Kenneth W. Dobie
;   TITLE OF INVENTION: ANTISENSE MODULATION OF PROX-1 EXPRESSION
;   FILE REFERENCE: RTS-0421
;   CURRENT APPLICATION NUMBER: US/10/162,846
;   CURRENT FILING DATE:  2002-06-03
;   NUMBER OF SEQ ID NOS: 134
; SEQ ID NO 11
;   LENGTH: 48396
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-10-162-846-11

Alignment Scores:
Pred. No.:             4.11e+03          Length:             48396
Score:                 9.00              Matches:            9
```

Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-162-846-11 (1-48396)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
          |||||||||||||||||||||||||||
Db   75 CTCCCAAGTTCTCTTGCCTTGCATCC 101

RESULT 90
US-10-471-010-3
; Sequence 3, Application US/10471010
; Publication No. US20040185527A1
; GENERAL INFORMATION:
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001044-US
; CURRENT APPLICATION NUMBER: US/10/471,010
; CURRENT FILING DATE: 2003-09-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 73544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(73544)
; OTHER INFORMATION: n = A,T,C or G
US-10-471-010-3

Alignment Scores:
Pred. No.:   5.84e+03   Length: 73544
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x US-10-471-010-3 (1-73544)

Qy   24 LeuProSerSerLeuAlaLeuLeuSer 32
          |||||||||||||||||||||||||||
Db 70413 TTGGCATCCCTCTTGCCTTGTTGTCC 70439

RESULT 91
US-09-997-722-85/c
; Sequence 85, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 85
; LENGTH: 96600
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7789)..(7808)

; OTHER INFORMATION: "n" at positions 7789 through 7808 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9978)..(9997)
; OTHER INFORMATION: "n" at positions 9978 through 9997 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (14125)..(14144)
; OTHER INFORMATION: "n" at positions 14125 through 14144 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (20780)..(20799)
; OTHER INFORMATION: "n" at positions 20780 through 20799 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (27797)..(27816)
; OTHER INFORMATION: "n" at positions 27797 through 27816 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (31149)..(31168)
; OTHER INFORMATION: "n" at positions 31149 through 31168 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (33786)..(33805)
; OTHER INFORMATION: "n" at positions 33786 through 33805 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34873)..(34892)
; OTHER INFORMATION: "n" at positions 34873 through 34892 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45961)..(45980)
; OTHER INFORMATION: "n" at positions 45961 through 45980 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (50121)..(50140)
; OTHER INFORMATION: "n" at positions 50121 through 50140 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (53877)..(53896)
; OTHER INFORMATION: "n" at positions 53877 through 53896 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55271)..(55290)
; OTHER INFORMATION: "n" at positions 55271 through 55290 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (57012)..(57031)
; OTHER INFORMATION: "n" at positions 57012 through 57031 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (60915)..(60934)
; OTHER INFORMATION: "n" at positions 60915 through 60934 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (63387)..(63406)
; OTHER INFORMATION: "n" at positions 63387 through 63406 can be any base.
US-09-997-722-85

Alignment Scores:
Pred. No.:   7.34e+03   Length: 96600
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 3   Gaps: 0

US-09-989-890-238 (1-212) x US-09-997-722-85 (1-96600)

Qy   54 LeuProSerAlaSerAlaAlaAlaGly 62
          |||||||||||||||||||||||||||
Db 29168 CTCCCTAGTGCCAGTGCTGCAGCTGGT 29142

RESULT 92
US-10-156-761-15103/c
; Sequence 15103, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15103
; LENGTH: 100000
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15103

Alignment Scores:
Pred. No.:   7.55e+03   Length: 100000
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-156-761-15103 (1-100000)

Qy   52 AlaGlySerSerAlaSerAlaAla 60
                 |||||||||||||||||||||||||
Db   75162 GCGGGTTTGCCGTCGGCGAGCGCCGCG 75136

RESULT 93
US-10-981-277-31/c
; Sequence 31, Application US/10981277
; Publication No. US20050181389A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa
; TITLE OF INVENTION: Compositions and Methods for Glioma Classification
; FILE REFERENCE: 03-968-US
; CURRENT APPLICATION NUMBER: US/10/981,277
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/516,817
; PRIOR FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 31
; LENGTH: 183046
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-981-277-31

Alignment Scores:
Pred. No.:   1.25e+04   Length: 183046
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 9   Gaps: 0

US-09-989-890-238 (1-212) x US-10-981-277-31 (1-183046)

Qy   22 ArgAlaLeuProSerSerLeuAlaLeu 30

Db   167188 AGAGCCCTTCCATCCTCCTTAGCCCTC 167162

RESULT 94
US-10-719-993-6787/c
; Sequence 6787, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6787
; LENGTH: 493999
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(493999)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
; ORGANISM: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6787

Alignment Scores:
Pred. No.:   2.88e+04   Length: 493999
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x US-10-719-993-6787 (1-493999)

Qy   127 ThrProMetLeuSerSerLysAlaSer 135
                 |||||||||||||||||||||||||
Db   222791 ACACCTATGCTCCTCCAAAGCTTCA 222765

RESULT 95
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.:   3.25e+05   Length: 9025608

Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-156-761-1 (1-9025608)

Qy   52 AlaGlyLeuProSerAlaSerAlaAla 60
              ||||||||||||||||||||||||||||
Db   502750 GCGGGTTTGCCGTCGGCGAGCGCCGCG 502776

RESULT 96
US-10-719-900-172651/c
; Sequence 172651, Application US/10719900
; Publication No. US2005002164A1
; GENERAL INFORMATION:
;   APPLICANT: Xue Mei Zhou
;   TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
;   FILE REFERENCE: 3528.1
;   CURRENT APPLICATION NUMBER: US/10/719,900
;   CURRENT FILING DATE: 2003-11-20
;   PRIOR APPLICATION NUMBER: 60/427,808
;   PRIOR FILING DATE: 2002 11 20
;   NUMBER OF SEQ ID NOS: 982914
;   SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 172651
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-719-900-172651

Alignment Scores:
Pred. No.:   Length: 25
Score: 65.1   Matches: 8
Percent Similarity: 8.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 3.8%   Gaps: 0
                      8

US-09-989-890-238 (1-212) x US-10-719-900-172651 (1-25)

Qy   115 SerGlySerSerGlnProTrpAla 122
              ||||||||||||||||||||||||
Db   24 TCCGGATCCTAGTCAGCCCCTGGGCT 1

RESULT 97
US-11-036-317-669570
; Sequence 669570, Application US/11036317
; Publication No. US2005021482A1
; GENERAL INFORMATION:
;   APPLICANT: Williams, Alan
;   APPLICANT: Blume, John
;   TITLE OF INVENTION: Method of Analysis of Alternative Splicing in Mouse
;   FILE REFERENCE: 3654.1
;   CURRENT APPLICATION NUMBER: US/11/036,317
;   CURRENT FILING DATE: 2005-01-13
;   PRIOR APPLICATION NUMBER: US 60/536,639
;   PRIOR FILING DATE: 2004-01-13
;   NUMBER OF SEQ ID NOS: 991174
;   SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 669570
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-11-036-317-669570

Alignment Scores:
Pred. No.:   Length: 25
Score: 65.1   Matches: 8
Percent Similarity: 8.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
                      3.8%

DB: 10   Gaps: 0

US-09-989-890-238 (1-212) x US-11-036-317-669570 (1-25)

Qy   129 MetLeuSerSerLysAlaSerLeu 136
              ||||||||||||||||||||||||
Db   2 ATGTTAAGCTCAAAGGCATCCCTG 25

RESULT 98
US-11-036-317-669571
; Sequence 669571, Application US/11036317
; Publication No. US2005021482A1
; GENERAL INFORMATION:
;   APPLICANT: Williams, Alan
;   APPLICANT: Blume, John
;   TITLE OF INVENTION: Method of Analysis of Alternative Splicing in Mouse
;   FILE REFERENCE: 3654.1
;   CURRENT APPLICATION NUMBER: US/11/036,317
;   CURRENT FILING DATE: 2005-01-13
;   PRIOR APPLICATION NUMBER: US 60/536,639
;   PRIOR FILING DATE: 2004-01-13
;   NUMBER OF SEQ ID NOS: 991174
;   SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 669571
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-11-036-317-669571

Alignment Scores:
Pred. No.:   Length: 25
Score: 65.1   Matches: 8
Percent Similarity: 8.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 3.8%   Gaps: 0
                      10

US-09-989-890-238 (1-212) x US-11-036-317-669571 (1-25)

Qy   129 MetLeuSerSerLysAlaSerLeu 136
              ||||||||||||||||||||||||
Db   2 ATGTTAAGCTCAAAGGCATCCCTG 25

RESULT 99
US-11-036-317-687056
; Sequence 687056, Application US/11036317
; Publication No. US2005021482A1
; GENERAL INFORMATION:
;   APPLICANT: Williams, Alan
;   APPLICANT: Blume, John
;   TITLE OF INVENTION: Method of Analysis of Alternative Splicing in Mouse
;   FILE REFERENCE: 3654.1
;   CURRENT APPLICATION NUMBER: US/11/036,317
;   CURRENT FILING DATE: 2005-01-13
;   PRIOR APPLICATION NUMBER: US 60/536,639
;   PRIOR FILING DATE: 2004-01-13
;   NUMBER OF SEQ ID NOS: 991174
;   SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 687056
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-11-036-317-687056

Alignment Scores:
Pred. No.:   Length: 25
Score: 65.1   Matches: 8
Percent Similarity: 8.00   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 100.0%   Indels: 0
DB: 3.8%   Gaps: 0
                      10

US-09-989-890-238 (1-212) x US-11-036-317-687056 (1-25)

Qy   29 AlaLeuLeuSerArgProLeuSer 36
               |||||||||||||||||||||||||
Db   1 GCGCTGCTGAGTAGGCCTCTCTCT 24

RESULT 100
US-10-131-827-69
; Sequence 69, Application US/10131827
; Publication No. US2004009479A1
; GENERAL INFORMATION:
; APPLICANT: Wohlgemuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 50661200000120
; CURRENT APPLICATION NUMBER: US/10/131,827
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69
;   LENGTH: 50
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-131-827-69

Alignment Scores:
Pred. No.:   116   Length: 50
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-10-131-827-69 (1-50)

Qy   25 ProSerSerLeuAlaLeuLeuSer 32
               ||||||||||||||||||||||||
Db   23 CCTTCAAGCCTAGCCCTTCTCTCA 46

Search completed: March 17, 2006, 08:41:47
Job time : 3281 secs

OM protein - nucleic search, using frame_plus_p2n model

                    Copyright (c) 1993 - 2006 Bioacceleration Ltd.

                          GenCore version 5.1.7

Run on: March 17, 2006, 06:38:36 ; Search time 1251 Seconds
                                          (without alignments)
                                          395.141 Million cell updates/sec

Title: US-09-989-890-238
Perfect score: 212
Sequence: 1 SPHQAAAPVDQTPRTLATMG.........RRWAVAPCRAEKLMCSSSRS 212

Scoring table: OLIGO
           Xgapop 60.0 , Xgapext 60.0
           Ygapop 60.0 , Ygapext 60.0
           Fgapop   6.0 , Fgapext   7.0
           Delop   6.0 , Delext   7.0

Searched: 8023312 seqs, 1165852854 residues

Word size: 1

Total number of hits satisfying chosen parameters: 16044706

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DBV=x1h
-Q=/abss/ABSSWEB_spool/US09989890/runat_16032006_095523_17709/app_query.fasta_1
-DB=Published_Applications_NA_New -QFMT=fastap -SUFFIX=oligo_p2n.rnpbn
-MINMATCH=0.1 -LOOPEXT=0 -LOOPCL=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=500 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=100 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abss05h
-USER=US09989890 @CGN_1_1_431 @runat_16032006_095523_17709 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA_New:*
        1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
        2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
        3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
        4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
        5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
        6: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
        7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
        8: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
        9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
       10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
       11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
       12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
       13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		115	1785	9		US-11-072-512-799	Sequence 799, App
2	c	54.2	201	12		US-11-124-368A-2250	Sequence 2250, Ap
3	c	4.2	201	12		US-11-124-368A-2364	Sequence 2364, Ap
4	c	4.2	201	12		US-11-124-368A-2475	Sequence 2475, Ap
5	c	9	4.2	201	12	US-11-124-368A-17847	Sequence 17847, A
6	c	9	4.2	201	12	US-11-124-367A-3389	Sequence 3389, Ap
7		9	4.2	201	12	US-11-124-367A-3491	Sequence 3491, Ap
8	c	9	4.2	201	12	US-11-124-367A-3594	Sequence 3594, Ap
9		9	4.2	201	12	US-11-124-367A-22366	Sequence 22366, A
10	c	9	4.2	637	6	US-09-925-065A-468689	Sequence 468689,
11	c	9	4.2	652	6	US-09-925-065A-726141	Sequence 726141,
12		9	4.2	718	8	US-10-750-185-48878	Sequence 48878, A
13		9	4.2	718	8	US-10-750-623-48878	Sequence 48878, A
14	c	9	4.2	1209	7	US-10-649-591-9	Sequence 9, Appli
15		9	4.2	2520	8	US-10-750-065A-691773	Sequence 691773,
16		9	4.2	2888	8	US-10-750-185-53796	Sequence 53796, A
17		9	4.2	2888	8	US-10-750-623-53796	Sequence 53796, A
18	c	9	4.2	11598	12	US-11-124-368A-137	Sequence 137, App
19	c	9	4.2	11598	12	US-11-124-367A-173	Sequence 173, App
20	c	9	4.2	11602	12	US-11-124-368A-136	Sequence 136, App
21	c	9	4.2	11602	12	US-11-124-367A-172	Sequence 172, App
22	c	9	4.2	12678	12	US-11-124-368A-135	Sequence 135, App
23	c	9	4.2	12678	12	US-11-124-367A-171	Sequence 171, App
24	c	9	4.2	41041	12	US-11-124-368A-2920	Sequence 2920, Ap
25	c	9	4.2	41041	12	US-11-124-367A-5068	Sequence 5068, Ap
26	c	9	4.2	215248	7	US-10-330-773-261	Sequence 261, App
27		8	3.8	25	12	US-11-136-527-359961	Sequence 359961,
28	c	8	3.8	198	12	US-11-103-122-44	Sequence 44, Appl
29		8	3.8	201	12	US-11-124-367A-24790	Sequence 24790, A
30		8	3.8	201	12	US-11-124-367A-24791	Sequence 24791, A
31		8	3.8	201	12	US-11-124-367A-25234	Sequence 25234, A
32	c	8	3.8	317	6	US-09-925-065A-620441	Sequence 620441,
33		8	3.8	377	6	US-09-925-065A-363884	Sequence 363884,
34		8	3.8	377	6	US-09-925-065A-363885	Sequence 363885,
35		8	3.8	377	6	US-09-925-065A-363886	Sequence 363886,
36	c	8	3.8	395	9	US-11-043-752-949	Sequence 949, App
37		8	3.8	406	9	US-11-000-688-608	Sequence 608, App
38	c	8	3.8	408	6	US-09-925-065A-844771	Sequence 844771,
39		8	3.8	441	8	US-10-623-155-97	Sequence 97, Appl
40		8	3.8	473	6	US-09-925-065A-489934	Sequence 489934,
41	c	8	3.8	475	7	US-10-667-295-223	Sequence 223, App
42	c	8	3.8	497	7	US-10-667-295-15	Sequence 15, Appl
43		8	3.8	517	6	US-09-925-065A-616675	Sequence 616675,
44	c	8	3.8	517	6	US-09-925-065A-616676	Sequence 616676,
45	c	8	3.8	518	6	US-09-925-065A-418469	Sequence 418469,
46		8	3.8	525	7	US-10-467-657-2511	Sequence 2511, Ap
47		8	3.8	542	9	US-11-021-492-486	Sequence 486, App
48		8	3.8	547	6	US-09-925-065A-151898	Sequence 151898,
49		8	3.8	552	6	US-09-925-065A-748081	Sequence 748081,
50	c	8	3.8	559	6	US-09-925-065A-808255	Sequence 808255,
51	c	8	3.8	559	6	US-09-925-065A-808256	Sequence 808256,
52	c	8	3.8	559	6	US-09-925-065A-808257	Sequence 808257,
53	c	8	3.8	565	6	US-09-925-065A-946604	Sequence 946604,
54		8	3.8	570	6	US-09-925-065A-693639	Sequence 693639,
55		8	3.8	570	6	US-09-925-065A-950741	Sequence 950741,
56	c	8	3.8	571	6	US-09-925-065A-733926	Sequence 733926,
57		8	3.8	576	6	US-09-925-065A-854936	Sequence 854936,
58		8	3.8	581	6	US-09-925-065A-53760	Sequence 53760, A
59	c	8	3.8	581	6	US-09-925-065A-53761	Sequence 53761, A
60	c	8	3.8	581	6	US-09-925-065A-53762	Sequence 53762, A
61	c	8	3.8	586	6	US-09-925-065A-295614	Sequence 295614,
62	c	8	3.8	593	6	US-09-925-065A-630739	Sequence 630739,
63		8	3.8	595	6	US-09-925-065A-295459	Sequence 295459,
64		8	3.8	595	6	US-09-925-065A-641425	Sequence 641425,
65	c	8	3.8	599	6	US-09-925-065A-301227	Sequence 301227,
66		8	3.8	599	6	US-09-925-065A-301228	Sequence 301228,
67		8	3.8	600	8	US-10-623-155-98	Sequence 98, Appl
68		8	3.8	602	12	US-11-136-527-8133	Sequence 8133, Ap
69	c	8	3.8	602	6	US-09-925-065A-774077	Sequence 774077,
70		8	3.8	603	6	US-09-925-065A-837176	Sequence 837176,
71	c	8	3.8	609	6	US-09-925-065A-571840	Sequence 571840,
72	c	8	3.8	611	6	US-09-925-065A-630748	Sequence 630748,
73		8	3.8	621	6	US-09-925-065A-893900	Sequence 893900,
74		8	3.8	622	6	US-09-925-065A-894636	Sequence 894636,
75	c	8	3.8	622	6	US-09-925-065A-464419	Sequence 464419,
76	c	8	3.8	623	6	US-09-925-065A-672678	Sequence 672678,
77	c	8	3.8	623	6	US-09-925-065A-686260	Sequence 686260,

c	78	8	3.8	624	US-09-925-065A-31084	6	Sequence 31084,
c	79	8	3.8	624	US-09-925-065A-749727	6	Sequence 749727,
c	80	8	3.8	624	US-09-925-065A-822330	6	Sequence 822330,
c	81	8	3.8	637	US-09-925-065A-883750	6	Sequence 883750,
	82	8	3.8	637	US-09-925-065A-910175	6	Sequence 910175,
c	83	8	3.8	639	US-09-925-065A-709912	6	Sequence 709912,
c	84	8	3.8	662	US-09-925-065A-441386	6	Sequence 441386,
c	85	8	3.8	664	US-09-925-065A-4577	6	Sequence 4577, Ap
c	86	8	3.8	664	US-09-925-065A-4578	6	Sequence 4578, Ap
c	87	8	3.8	664	US-09-925-065A-4579	6	Sequence 4579, Ap
c	88	8	3.8	710	US-11-103-957-91	12	Sequence 91, Appl
c	89	8	3.8	829	US-11-077-619-41	9	Sequence 41, Appl
c	90	8	3.8	855	US-10-821-234-780	9	Sequence 780, App
c	91	8	3.8	1002	US-11-128-061-61	12	Sequence 61, Appl
c	92	8	3.8	1002	US-11-128-049-61	12	Sequence 61, Appl
c	93	8	3.8	1038	US-09-925-065A-63837	6	Sequence 63837, A
c	94	8	3.8	1038	US-09-925-065A-63838	6	Sequence 63838, A
c	95	8	3.8	1038	US-09-925-065A-63839	6	Sequence 63839, A
c	96	8	3.8	1044	US-11-091-883-130	12	Sequence 130, App
c	97	8	3.8	1056	US-09-925-065A-576	6	Sequence 576, App
c	98	8	3.8	1056	US-09-925-065A-577	6	Sequence 577, App
c	99	8	3.8	1056	US-09-925-065A-578	6	Sequence 578, App
c	100	8	3.8	1056	US-09-925-065A-579	6	Sequence 579, App
c	101	8	3.8	1187	US-09-925-065A-279504	6	Sequence 279504,
c	102	8	3.8	1220	US-10-750-185-30885	8	Sequence 30885, A
c	103	8	3.8	1220	US-10-750-185-59964	8	Sequence 59964, A
c	104	8	3.8	1220	US-10-750-622-30885	8	Sequence 30885, A
c	105	8	3.8	1220	US-10-750-623-59964	8	Sequence 59964, A
c	106	8	3.8	1304	US-11-112-908-457	12	Sequence 457, App
c	107	8	3.8	1358	US-10-750-185-59809	8	Sequence 59809, A
c	108	8	3.8	1358	US-10-750-623-59809	8	Sequence 59809, A
c	109	8	3.8	1400	US-11-128-061-4184	12	Sequence 4184, Ap
c	110	8	3.8	1400	US-11-128-049-4184	12	Sequence 4184, Ap
c	111	8	3.8	1430	US-10-750-185-61160	8	Sequence 61160, A
c	112	8	3.8	1430	US-10-750-623-61160	8	Sequence 61160, A
c	113	8	3.8	1465	US-09-925-065A-713516	6	Sequence 713516,
c	114	8	3.8	1538	US-11-096-568A-9080	9	Sequence 9080, Ap
c	115	8	3.8	1611	US-10-750-185-37511	8	Sequence 37511, A
c	116	8	3.8	1611	US-10-750-622-37511	8	Sequence 37511, A
c	117	8	3.8	1634	US-10-750-185-43444	8	Sequence 43444, A
c	118	8	3.8	1650	US-10-932-182A-77528	7	Sequence 77528, A
c	119	8	3.8	1650	US-10-750-185-77528	8	Sequence 77528, A
c	120	8	3.8	1672	US-11-096-568A-15814	9	Sequence 15814, A
c	121	8	3.8	1685	US-10-750-185-64640	8	Sequence 64640, A
c	122	8	3.8	1685	US-10-750-622-64640	8	Sequence 64640, A
c	123	8	3.8	1743	US-09-925-065A-707466	6	Sequence 707466,
c	124	8	3.8	1743	US-09-925-065A-707467	6	Sequence 707467,
c	125	8	3.8	1775	US-11-136-527-4037	12	Sequence 4037, Ap
c	126	8	3.8	1780	US-10-750-185-25811	8	Sequence 25811, A
c	127	8	3.8	1780	US-10-750-623-25811	8	Sequence 25811, A
c	128	8	3.8	1784	US-11-096-568A-18800	9	Sequence 18800, Ap
c	129	8	3.8	1837	US-10-750-185-38734	8	Sequence 38734, A
c	130	8	3.8	1837	US-10-750-623-38734	8	Sequence 38734, A
c	131	8	3.8	2027	US-09-925-065A-706432	6	Sequence 706432,
c	132	8	3.8	2342	US-10-821-234-199	9	Sequence 199, Appli
c	133	8	3.8	2342	US-10-507-755-2	7	Sequence 2, Appli
c	134	8	3.8	2455	US-11-122-329-46	12	Sequence 46, Appl
c	135	8	3.8	2499	US-11-467-657-3875	9	Sequence 3875, Ap
c	136	8	3.8	2552	US-11-072-512-516	9	Sequence 516, App
c	137	8	3.8	2914	US-11-072-512-1009	12	Sequence 1009, Ap
c	138	8	3.8	3387	US-11-219-611-2	12	Sequence 2, Appli
c	139	8	3.8	3407	US-11-000-688-89	12	Sequence 89, Appli
c	140	8	3.8	3498	US-11-072-512-952	9	Sequence 952, App
c	141	8	3.8	3584	US-11-072-512-538	9	Sequence 538, App
c	142	8	3.8	3868	US-10-501-035-131	7	Sequence 131, App
c	143	8	3.8	3899	US-11-128-061-542	12	Sequence 542, App
c	144	8	3.8	3899	US-11-128-049-542	12	Sequence 542, App
c	145	8	3.8	4406	US-11-072-512-979	9	Sequence 979, App
c	146	8	3.8	4797	US-10-623-155-134	8	Sequence 134, App
c	147	8	3.8	5220	US-11-110-204-1	12	Sequence 1, Appli
c	148	8	3.8	19691	US-11-124-367A-5034	12	Sequence 5034, Ap
c	149	8	3.8	37898	US-10-330-773-759	7	Sequence 759, App
c	150	8	3.8	39762	US-11-124-367A-5039	12	Sequence 5039, Ap

c	151	8	3.8	88892	7	US-10-330-773-672	Sequence 672, App
c	152	8	3.8	95604	12	US-11-124-367A-5097	Sequence 5097, Ap
c	153	8	3.8	100000	12	US-11-124-367A-5086	Sequence 5086, Ap
c	154	8	3.8	151828	12	US-11-117-187-197	Sequence 197, App
c	155	8	3.8	151870	8	US-10-995-561-13199	Sequence 13199, A
c	156	8	3.8	155515	12	US-11-112-908-42	Sequence 42, Appl
c	157	8	3.8	159660	12	US-11-112-908-43	Sequence 43, Appl
	158	8	3.8	162537	12	US-11-121-086-59	Sequence 59, Appl
	159	8	3.8	170837	12	US-11-121-086-97	Sequence 97, Appl
	160	8	3.8	177623	12	US-11-112-908-41	Sequence 41, Appl
	161	8	3.8	179666	12	US-11-121-086-67	Sequence 67, Appl
	162	8	3.8	179777	12	US-11-121-086-106	Sequence 106, App
	163	8	3.8	185750	9	US-11-114-798-56	Sequence 56, Appl
	164	8	3.8	195998	8	US-10-995-561-13489	Sequence 13489, A
	165	8	3.8	215974	8	US-10-330-773-500	Sequence 500, App
	166	8	3.8	228006	7	US-10-330-773-119	Sequence 119, App
	167	8	3.8	390183	7	US-10-330-773-937	Sequence 937, App
c	168	8	3.8	1080000	8	US-10-928-446A-1	Sequence 1, Appli
c	169	8	3.8	1080000	8	US-10-928-446A-181	Sequence 181, App
c	170	8	3.8	1080000	8	US-10-928-446A-183	Sequence 183, App
c	171	8	3.8	1080000	8	US-10-928-446A-185	Sequence 185, App
c	172	8	3.8	1080000	8	US-10-928-446A-187	Sequence 187, App
c	173	8	3.8	1080000	8	US-10-928-446A-189	Sequence 189, App
c	174	8	3.8	1080000	8	US-10-928-446A-191	Sequence 191, App
c	175	8	3.8	1080000	8	US-10-928-446A-193	Sequence 193, App
c	176	8	3.8	1080000	8	US-10-928-446A-195	Sequence 195, App
c	177	8	3.8	1080000	8	US-10-928-446A-197	Sequence 197, App
c	178	8	3.8	1080000	8	US-10-928-446A-199	Sequence 199, App
c	179	8	3.8	1080000	8	US-10-928-446A-201	Sequence 201, App
c	180	7	3.3	21	8	US-10-310-914A-97586	Sequence 97586, A
c	181	7	3.3	21	8	US-10-310-914A-97587	Sequence 97587, A
c	182	7	3.3	21	7	US-10-310-914A-101787	Sequence 101787,
c	183	7	3.3	21	8	US-10-310-914A-136191	Sequence 136191,
c	184	7	3.3	21	8	US-10-310-914A-279267	Sequence 279267,
c	185	7	3.3	21	8	US-10-310-914A-383155	Sequence 383155,
c	186	7	3.3	22	8	US-10-310-914A-87973	Sequence 87973, A
c	187	7	3.3	22	8	US-10-310-914A-96746	Sequence 96746, A
c	188	7	3.3	22	8	US-10-310-914A-101811	Sequence 101811,
c	189	7	3.3	22	8	US-10-310-914A-215941	Sequence 215941,
c	190	7	3.3	22	8	US-10-310-914A-324578	Sequence 324578,
c	191	7	3.3	22	8	US-10-310-914A-361254	Sequence 361254,
c	192	7	3.3	23	8	US-10-310-914A-372019	Sequence 372019,
c	193	7	3.3	22	8	US-10-310-914A-877785	Sequence 877785,
c	194	7	3.3	23	8	US-10-310-914A-1112628	Sequence 1112628,
c	195	7	3.3	23	8	US-10-310-914A-152009	Sequence 152009,
c	196	7	3.3	23	8	US-10-310-914A-372045	Sequence 372045,
c	197	7	3.3	23	8	US-10-310-914A-599594	Sequence 599594,
c	198	7	3.3	24	8	US-10-310-914A-38166	Sequence 38166, A
c	199	7	3.3	24	8	US-10-310-914A-148005	Sequence 148005,
c	200	7	3.3	24	8	US-10-310-914A-862534	Sequence 862534,
c	201	7	3.3	24	8	US-10-310-914A-1112531	Sequence 1112531,
c	202	7	3.3	24	8	US-10-310-914A-1224066	Sequence 1224066,
c	203	7	3.3	25	7	US-10-932-182A-69608	Sequence 69608, A
c	204	7	3.3	25	7	US-10-932-182A-94852	Sequence 94852, A
c	205	7	3.3	25	7	US-10-932-182A-69608	Sequence 69608, A
c	206	7	3.3	25	7	US-10-932-182A-94852	Sequence 94852, A
c	207	7	3.3	25	7	US-10-933-982-57548	Sequence 57548, A
c	208	7	3.3	25	7	US-10-933-982-57553	Sequence 57553, A
c	209	7	3.3	25	7	US-10-933-982-57554	Sequence 57554, A
c	210	7	3.3	25	7	US-10-933-982-57555	Sequence 57555, A
c	211	7	3.3	25	7	US-10-933-982-57560	Sequence 57560, A
c	212	7	3.3	25	7	US-10-933-982-57564	Sequence 57564, A
c	213	7	3.3	25	12	US-11-121-849-182011	Sequence 182011,
c	214	7	3.3	25	12	US-11-121-849-204560	Sequence 204560,
c	215	7	3.3	25	7	US-10-934-048A-60292	Sequence 60292, A
c	216	7	3.3	25	12	US-11-121-849-85832	Sequence 85832, A
c	217	7	3.3	25	12	US-11-121-849-213336	Sequence 213336,
c	218	7	3.3	25	12	US-11-121-849-530519	Sequence 204560,
c	219	7	3.3	25	12	US-11-121-849-530519	Sequence 530519,
c	220	7	3.3	25	12	US-11-136-527-254296	Sequence 254296,
c	221	7	3.3	25	12	US-11-136-527-289028	Sequence 289028,
c	222	7	3.3	25	12	US-11-136-527-346564	Sequence 346564,
c	223	7	3.3	26	8	US-10-310-914A-96747	Sequence 96747, A

Index	c			N1	N2	Reference	Sequence
224	c	3.3	7	26	8	US-10-310-914A-101773	Sequence 101773,
225	c	3.3	7	26	8	US-10-310-914A-877786	Sequence 877786,
226	c	3.3	7	26	8	US-10-310-914A-918187	Sequence 918187,
227	c	3.3	7	29	9	US-11-241-035-5	Sequence 5, Appli
228	c	3.3	7	32	8	US-10-939-294A-16780	Sequence 16780, A
229	c	3.3	7	32	8	US-10-939-294A-19543	Sequence 19543, A
230	c	3.3	7	50	12	US-11-175-859-21843	Sequence 21843, A
231	c	3.3	7	50	8	US-11-175-859-46314	Sequence 46314, A
232	c	3.3	7	50	12	US-11-175-859-104231	Sequence 104231,
233	c	3.3	7	61	8	US-10-310-914A-14090	Sequence 14090, A
234	c	3.3	7	66	8	US-10-310-914A-17565	Sequence 17565, A
235		3.3	7	77	8	US-10-310-914A-7601	Sequence 7601, Ap
236	c	3.3	7	84	8	US-10-310-914A-17562	Sequence 17562, A
237	c	3.3	7	88	8	US-10-310-914A-6599	Sequence 6599, Ap
238	c	3.3	7	89	8	US-10-310-914A-4425	Sequence 4425, Ap
239	c	3.3	7	96	8	US-10-310-914A-6615	Sequence 6615, Ap
240	c	3.3	7	108	8	US-10-310-914A-15226	Sequence 15226, A
241	c	3.3	7	108	8	US-10-310-914A-18377	Sequence 18377, A
242	c	3.3	7	180	8	US-11-167-614-7	Sequence 7, Appli
243	c	3.3	7	200	12	US-11-098-686-4071	Sequence 4071, Ap
244	c	3.3	7	201	12	US-10-995-561-1381	Sequence 1381, Ap
245	c	3.3	7	201	8	US-10-995-561-1432	Sequence 1432, Ap
246	c	3.3	7	201	8	US-10-995-561-1483	Sequence 1483, Ap
247	c	3.3	7	201	8	US-10-995-561-3436	Sequence 3436, Ap
248	c	3.3	7	201	8	US-10-995-561-3457	Sequence 3457, Ap
249	c	3.3	7	201	8	US-10-995-561-9503	Sequence 9503, Ap
250	c	3.3	7	201	8	US-10-995-561-9519	Sequence 9519, Ap
251	c	3.3	7	201	8	US-10-995-561-11235	Sequence 11235, A
252	c	3.3	7	201	8	US-10-995-561-11245	Sequence 11245, A
253	c	3.3	7	201	8	US-10-995-561-13035	Sequence 13035, A
254	c	3.3	7	201	8	US-10-995-561-15474	Sequence 15474, A
255	c	3.3	7	201	8	US-10-995-561-16314	Sequence 16314, A
256	c	3.3	7	201	8	US-10-995-561-22654	Sequence 22654, A
257	c	3.3	7	201	8	US-10-995-561-26049	Sequence 26049, A
258	c	3.3	7	201	8	US-10-995-561-26082	Sequence 26082, A
259	c	3.3	7	201	8	US-10-995-561-31702	Sequence 31702, A
260	c	3.3	7	201	8	US-10-995-561-36241	Sequence 36241, A
261	c	3.3	7	201	8	US-10-995-561-36778	Sequence 36778, A
262	c	3.3	7	201	8	US-10-995-561-40810	Sequence 40810, A
263	c	3.3	7	201	8	US-10-995-561-40811	Sequence 40811, A
264	c	3.3	7	201	8	US-10-995-561-41086	Sequence 41086, A
265	c	3.3	7	201	8	US-10-995-561-42002	Sequence 42002, A
266	c	3.3	7	201	8	US-10-995-561-49580	Sequence 49580, A
267	c	3.3	7	201	8	US-10-995-561-49598	Sequence 49598, A
268	c	3.3	7	201	8	US-10-995-561-49624	Sequence 49624, A
269	c	3.3	7	201	8	US-10-995-561-51200	Sequence 51200, A
270	c	3.3	7	201	8	US-10-995-561-54681	Sequence 54681, A
271	c	3.3	7	201	8	US-10-995-561-58666	Sequence 58666, A
272	c	3.3	7	201	8	US-10-995-561-58693	Sequence 58693, A
273	c	3.3	7	201	8	US-10-995-561-59955	Sequence 59955, A
274	c	3.3	7	201	8	US-10-995-561-61555	Sequence 61555, A
275	c	3.3	7	201	8	US-10-995-561-61913	Sequence 61913, A
276	c	3.3	7	201	8	US-10-995-561-61914	Sequence 61914, A
277	c	3.3	7	201	8	US-10-995-561-61917	Sequence 61917, A
278	c	3.3	7	201	8	US-10-995-561-62667	Sequence 62667, A
279	c	3.3	7	201	8	US-10-995-561-62672	Sequence 62672, A
280	c	3.3	7	201	8	US-10-995-561-62774	Sequence 62774, A
281	c	3.3	7	201	8	US-10-995-561-62779	Sequence 62779, A
282	c	3.3	7	201	8	US-10-995-561-63217	Sequence 63217, A
283	c	3.3	7	201	8	US-10-995-561-63284	Sequence 63284, A
284	c	3.3	7	201	8	US-10-995-561-63452	Sequence 63452, A
285	c	3.3	7	201	8	US-10-995-561-64276	Sequence 64276, A
286	c	3.3	7	201	8	US-10-995-561-65827	Sequence 65827, A
287	c	3.3	7	201	8	US-10-995-561-66403	Sequence 66403, A
288	c	3.3	7	201	8	US-10-995-561-66809	Sequence 66809, A
289	c	3.3	7	201	8	US-10-995-561-66836	Sequence 66836, A
290	c	3.3	7	201	8	US-10-995-561-66863	Sequence 66863, A
291	c	3.3	7	201	8	US-10-995-561-68113	Sequence 68113, A
292	c	3.3	7	201	8	US-10-995-561-68115	Sequence 68115, A
293	c	3.3	7	201	8	US-10-995-561-68137	Sequence 68137, A
294	c	3.3	7	201	8	US-10-995-561-68994	Sequence 68994, A
295	c	3.3	7	201	8	US-10-995-561-69010	Sequence 69010, A
296	c	3.3	7	201	8	US-10-995-561-69584	Sequence 69584, A
297	c	3.3	7	201	8	US-10-995-561-69608	Sequence 69608, A
298	c	3.3	7	201	8	US-10-995-561-69634	Sequence 69634, A
299	c	3.3	7	201	8	US-10-995-561-75984	Sequence 75984, A
300		3.3	7	201	8	US-10-995-561-80758	Sequence 80758, A
301		3.3	7	201	8	US-10-995-561-82635	Sequence 82635, A
302		3.3	7	201	8	US-10-995-561-82663	Sequence 82663, A
303		3.3	7	201	8	US-10-995-561-84047	Sequence 84047, A
304		3.3	7	201	8	US-10-995-561-84164	Sequence 84164, A
305	c	3.3	7	201	8	US-10-995-561-84968	Sequence 84968, A
306		3.3	7	201	12	US-11-124-368A-6513	Sequence 6513, Ap
307		3.3	7	201	12	US-11-124-368A-6971	Sequence 6971, Ap
308		3.3	7	201	12	US-11-124-368A-13981	Sequence 13981, A
309		3.3	7	201	12	US-11-124-368A-15038	Sequence 15038, A
310		3.3	7	201	12	US-11-124-367A-1689	Sequence 1689, Ap
311		3.3	7	201	12	US-11-124-367A-4602	Sequence 4602, Ap
312		3.3	7	201	12	US-11-124-367A-4620	Sequence 4620, Ap
313		3.3	7	201	12	US-11-124-367A-8410	Sequence 8410, Ap
314		3.3	7	201	12	US-11-124-367A-843	Sequence 843, Ap
315	c	3.3	7	201	12	US-11-124-367A-11354	Sequence 11354, A
316		3.3	7	201	12	US-11-124-367A-14339	Sequence 14339, A
317		3.3	7	201	12	US-11-124-367A-14499	Sequence 14499, A
318		3.3	7	201	12	US-11-124-367A-15720	Sequence 15720, A
319		3.3	7	201	12	US-11-124-367A-18457	Sequence 18457, A
320		3.3	7	201	12	US-11-124-367A-18624	Sequence 18624, A
321		3.3	7	201	12	US-11-124-367A-18626	Sequence 18626, A
322		3.3	7	201	12	US-11-124-367A-26223	Sequence 26223, A
323		3.3	7	201	12	US-11-124-367A-26540	Sequence 26540, A
324		3.3	7	201	12	US-11-124-367A-26541	Sequence 26541, A
325		3.3	7	201	12	US-11-124-367A-26607	Sequence 26607, A
326		3.3	7	201	12	US-11-124-367A-26750	Sequence 26750, A
327		3.3	7	201	12	US-11-124-367A-29694	Sequence 29694, A
328		3.3	7	201	12	US-11-124-367A-29695	Sequence 29695, A
329		3.3	7	201	12	US-11-124-367A-31424	Sequence 31424, A
330		3.3	7	201	12	US-11-124-367A-32221	Sequence 32221, A
331		3.3	7	247	12	US-11-124-367A-328710	Sequence 328710,
332		3.3	7	280	6	US-09-925-065A-583911	Sequence 583911,
333		3.3	7	280	6	US-09-925-065A-583912	Sequence 583912,
334		3.3	7	291	8	US-10-467-657-427	Sequence 427, App
335		3.3	7	311	6	US-09-925-065A-230017	Sequence 230017,
336		3.3	7	311	6	US-09-925-065A-230018	Sequence 230018,
337		3.3	7	311	6	US-09-925-065A-230019	Sequence 230019,
338		3.3	7	322	6	US-09-925-065A-560441	Sequence 560441,
339		3.3	7	326	6	US-09-925-065A-487640	Sequence 487640,
340		3.3	7	326	6	US-09-925-065A-487641	Sequence 487641,
341		3.3	7	326	6	US-09-925-065A-487642	Sequence 487642,
342		3.3	7	331	6	US-09-925-065A-851081	Sequence 851081,
343		3.3	7	335	6	US-09-925-065A-844897	Sequence 844897,
344		3.3	7	350	6	US-09-925-065A-475418	Sequence 475418,
345		3.3	7	355	6	US-09-925-065A-524778	Sequence 524778,
346		3.3	7	366	8	US-10-467-657-605	Sequence 605, App
347		3.3	7	369	8	US-10-821-234-786	Sequence 786, App
348		3.3	7	373	6	US-09-925-065A-208059	Sequence 208059,
349		3.3	7	376	6	US-09-925-065A-853117	Sequence 853117,
350		3.3	7	379	6	US-09-925-065A-437281	Sequence 437281,
351		3.3	7	379	6	US-09-925-065A-437287	Sequence 437287,
352	c	3.3	7	381	8	US-10-467-657-585	Sequence 585, App
353	c	3.3	7	381	8	US-10-623-155-215	Sequence 215, App
354	c	3.3	7	385	6	US-09-925-065A-211991	Sequence 211991,
355		3.3	7	404	8	US-10-821-234-11017	Sequence 11017, A
356		3.3	7	404	6	US-09-925-065A-11018	Sequence 11018, A
357		3.3	7	407	6	US-09-925-065A-515758	Sequence 515758,
358		3.3	7	408	8	US-10-821-234-93	Sequence 93, Appl
359		3.3	7	408	8	US-10-821-234-718	Sequence 718, App
360		3.3	7	409	12	US-09-925-065A-337508	Sequence 337508,
361	c	3.3	7	411	12	US-11-043-752-244	Sequence 244, App
362	c	3.3	7	419	6	US-09-925-065A-204998	Sequence 204998,
363	c	3.3	7	419	6	US-09-925-065A-204999	Sequence 204999,
364	c	3.3	7	419	6	US-09-925-065A-205004	Sequence 205004,
365	c	3.3	7	419	6	US-09-925-065A-205005	Sequence 205005,
366	c	3.3	7	420	6	US-09-925-065A-286319	Sequence 286319,
367	c	3.3	7	420	6	US-09-925-065A-286320	Sequence 286320,
368	c	3.3	7	423	6	US-09-925-065A-409101	Sequence 409101,
369	c	3.3	7	424	6	US-09-925-065A-169196	Sequence 169196,

Left listing:

	No						Sequence ID	Sequence
c	370	7	3.3	428	6		US-09-925-065A-657398	Sequence 657398,
	371	7	3.3	431	6		US-09-925-065A-796238	Sequence 796238,
c	372	7	3.3	432	6		US-09-925-065A-343867	Sequence 343867,
c	373	7	3.3	432	6		US-09-925-065A-420827	Sequence 420827,
	374	7	3.3	432	6		US-09-925-065A-652796	Sequence 652796,
	375	7	3.3	438	12		US-09-925-065A-321203	Sequence 321203,
	376	7	3.3	438	12		US-11-128-061-2931	Sequence 2931, Ap
	377	7	3.3	438	12		US-11-128-061-6573	Sequence 6573, Ap
	378	7	3.3	438	12		US-11-128-049-2931	Sequence 2931, Ap
	379	7	3.3	438	12		US-11-128-049-6573	Sequence 6573, Ap
	380	7	3.3	439	8		US-10-821-234-374	Sequence 374, App
	381	7	3.3	445	6		US-09-925-065A-229542	Sequence 229542,
	382	7	3.3	445	6		US-09-925-065A-845039	Sequence 845039,
	383	7	3.3	446	6		US-09-925-065A-134800	Sequence 134800,
	384	7	3.3	447	6		US-09-925-065A-633193	Sequence 633193,
	385	7	3.3	447	6		US-09-925-065A-633194	Sequence 633194,
c	386	7	3.3	449	6		US-09-925-065A-382753	Sequence 382753,
c	387	7	3.3	452	6		US-09-925-065A-353166	Sequence 353166,
c	388	7	3.3	452	6		US-09-925-065A-353167	Sequence 353167,
c	389	7	3.3	452	6		US-09-925-065A-353169	Sequence 353169,
c	390	7	3.3	453	6		US-09-925-065A-492024	Sequence 492024,
c	391	7	3.3	453	6		US-09-925-065A-492025	Sequence 492025,
c	392	7	3.3	453	6		US-09-925-065A-493433	Sequence 493433,
c	393	7	3.3	455	6		US-09-925-065A-133001	Sequence 133001,
c	394	7	3.3	455	6		US-09-925-065A-133002	Sequence 133002,
c	395	7	3.3	456	6		US-09-925-065A-413645	Sequence 413645,
c	396	7	3.3	457	6		US-09-925-065A-75971	Sequence 75971, A
c	397	7	3.3	458	6		US-09-925-065A-832468	Sequence 832468,
c	398	7	3.3	459	6		US-09-925-065A-542485	Sequence 542485,
c	399	7	3.3	459	6		US-09-925-065A-542486	Sequence 542486,
c	400	7	3.3	461	6		US-09-925-065A-365588	Sequence 365588,
c	401	7	3.3	461	6		US-09-925-065A-465773	Sequence 465773,
c	402	7	3.3	461	6		US-09-925-065A-465774	Sequence 465774,
c	403	7	3.3	462	6		US-09-925-065A-120956	Sequence 120956,
c	404	7	3.3	463	6		US-09-925-065A-525009	Sequence 525009,
c	405	7	3.3	463	12		US-11-112-908-356	Sequence 356, App
c	406	7	3.3	464	6		US-09-925-065A-543857	Sequence 543857,
c	407	7	3.3	468	8		US-10-821-234-401	Sequence 401, App
c	408	7	3.3	469	6		US-09-925-065A-358352	Sequence 358352,
c	409	7	3.3	469	6		US-09-925-065A-358353	Sequence 358353,
c	410	7	3.3	469	6		US-09-925-065A-358354	Sequence 358354,
c	411	7	3.3	469	6		US-09-925-065A-358355	Sequence 358355,
c	412	7	3.3	470	6		US-09-925-065A-495216	Sequence 495216,
c	413	7	3.3	471	6		US-09-925-065A-790129	Sequence 790129,
c	414	7	3.3	473	6		US-09-925-065A-633322	Sequence 633322,
c	415	7	3.3	477	6		US-09-925-065A-104340	Sequence 104340,
c	416	7	3.3	478	6		US-09-925-065A-201177	Sequence 201177,
c	417	7	3.3	478	6		US-09-925-065A-252979	Sequence 252979,
c	418	7	3.3	479	6		US-09-925-065A-404204	Sequence 404204,
c	419	7	3.3	479	6		US-09-925-065A-460322	Sequence 460322,
c	420	7	3.3	479	6		US-09-925-065A-625544	Sequence 625544,
c	421	7	3.3	480	6		US-09-925-065A-275127	Sequence 275127,
c	422	7	3.3	480	6		US-09-925-065A-783023	Sequence 783023,
c	423	7	3.3	480	6		US-09-925-065A-852992	Sequence 852992,
c	424	7	3.3	481	6		US-09-925-065A-155607	Sequence 155607,
c	425	7	3.3	481	6		US-09-925-065A-588853	Sequence 588853,
c	426	7	3.3	482	6		US-09-925-065A-35521	Sequence 35521, A
c	427	7	3.3	483	6		US-09-925-065A-311695	Sequence 311695,
c	428	7	3.3	483	6		US-09-925-065A-311696	Sequence 311696,
c	429	7	3.3	483	6		US-09-925-065A-311697	Sequence 311697,
c	430	7	3.3	483	6		US-09-925-065A-881354	Sequence 881354,
c	431	7	3.3	483	8		US-10-467-657-1337	Sequence 1337, Ap
c	432	7	3.3	486	6		US-09-925-065A-144188	Sequence 144188,
c	433	7	3.3	486	6		US-09-925-065A-144189	Sequence 144189,
c	434	7	3.3	486	6		US-09-925-065A-295737	Sequence 295737,
c	435	7	3.3	488	6		US-09-925-065A-127695	Sequence 127695,
c	436	7	3.3	488	6		US-09-925-065A-238234	Sequence 238234,
c	437	7	3.3	488	6		US-09-925-065A-238235	Sequence 238235,
c	438	7	3.3	489	6		US-09-925-065A-510758	Sequence 510758,
c	439	7	3.3	489	6		US-09-925-065A-322228	Sequence 322228,
c	440	7	3.3	490	6		US-09-925-065A-468074	Sequence 468074,
c	441	7	3.3	491	6		US-09-925-065A-504380	Sequence 504380,
c	442	7	3.3	491	6		US-09-925-065A-504381	Sequence 504381,

Right listing:

	No	Len			Pos	Sequence ID	Sequence
c	443	6	7	3.3	498	US-09-925-065A-148074	Sequence 148074,
	444	6	7	3.3	500	US-09-925-065A-574192	Sequence 574192,
	445	6	7	3.3	500	US-09-925-065A-574193	Sequence 574193,
c	446	6	7	3.3	501	US-09-925-065A-517675	Sequence 517675,
c	447	12	7	3.3	501	US-11-132-722-60	Sequence 60, Appl
c	448	12	7	3.3	501	US-11-132-722-61	Sequence 61, Appl
c	449	12	7	3.3	501	US-11-132-722-63	Sequence 63, Appl
	450	12	7	3.3	501	US-11-132-722-64	Sequence 64, Appl
c	451	12	7	3.3	501	US-11-132-722-66	Sequence 66, Appl
c	452	12	7	3.3	501	US-11-132-722-67	Sequence 67, Appl
c	453	12	7	3.3	501	US-11-132-722-69	Sequence 69, Appl
	454	12	7	3.3	501	US-11-132-722-70	Sequence 70, Appl
	455	12	7	3.3	501	US-11-132-722-72	Sequence 72, Appl
	456	12	7	3.3	501	US-11-132-722-73	Sequence 73, Appl
	457	12	7	3.3	501	US-11-132-722-75	Sequence 75, Appl
	458	12	7	3.3	501	US-11-132-722-76	Sequence 76, Appl
	459	6	7	3.3	504	US-09-925-065A-225667	Sequence 225667,
	460	6	7	3.3	504	US-09-925-065A-225668	Sequence 225668,
c	461	6	7	3.3	504	US-09-925-065A-784979	Sequence 784979,
c	462	6	7	3.3	504	US-09-925-065A-844704	Sequence 844704,
c	463	6	7	3.3	505	US-09-925-065A-237346	Sequence 237346,
	464	6	7	3.3	505	US-09-925-065A-237347	Sequence 237347,
	465	6	7	3.3	505	US-09-925-065A-237348	Sequence 237348,
	466	6	7	3.3	505	US-09-925-065A-237349	Sequence 237349,
	467	6	7	3.3	505	US-09-925-065A-480918	Sequence 480918,
	468	6	7	3.3	505	US-09-925-065A-480919	Sequence 480919,
c	469	6	7	3.3	505	US-09-925-065A-480920	Sequence 480920,
c	470	6	7	3.3	505	US-09-925-065A-480921	Sequence 480921,
	471	6	7	3.3	507	US-09-925-065A-794810	Sequence 794810,
	472	6	7	3.3	507	US-09-925-065A-228166	Sequence 228166,
c	473	6	7	3.3	508	US-09-925-065A-22479	Sequence 22479, A
	474	6	7	3.3	508	US-09-925-065A-398214	Sequence 398214,
c	475	6	7	3.3	508	US-09-925-065A-398215	Sequence 398215,
	476	6	7	3.3	509	US-09-925-065A-255117	Sequence 255117,
c	477	6	7	3.3	509	US-09-925-065A-767243	Sequence 767243,
c	478	6	7	3.3	509	US-09-925-065A-830574	Sequence 830574,
c	479	6	7	3.3	510	US-09-925-065A-327180	Sequence 327180,
c	480	6	7	3.3	510	US-09-925-065A-327181	Sequence 327181,
c	481	6	7	3.3	511	US-09-925-065A-525587	Sequence 525587,
c	482	6	7	3.3	511	US-09-925-065A-525588	Sequence 525588,
c	483	7	7	3.3	511	US-10-475-075-434	Sequence 434, App
c	484	6	7	3.3	512	US-09-925-065A-765838	Sequence 765838,
c	485	6	7	3.3	512	US-09-925-065A-832567	Sequence 832567,
c	486	7	7	3.3	512	US-10-475-075-425	Sequence 425, App
c	487	6	7	3.3	513	US-09-925-065A-92984	Sequence 92984, A
c	488	6	7	3.3	513	US-09-925-065A-92985	Sequence 92985, A
c	489	6	7	3.3	513	US-09-925-065A-241387	Sequence 241387,
c	490	6	7	3.3	513	US-09-925-065A-324263	Sequence 324263,
c	491	6	7	3.3	513	US-09-925-065A-324264	Sequence 324264,
c	492	6	7	3.3	513	US-09-925-065A-774726	Sequence 774726,
c	493	6	7	3.3	513	US-09-925-065A-774727	Sequence 774727,
c	494	12	7	3.3	513	US-11-128-061-2342	Sequence 2342, Ap
	495	12	7	3.3	513	US-11-128-061-5984	Sequence 5984, Ap
	496	12	7	3.3	513	US-11-128-049-2342	Sequence 2342, Ap
	497	12	7	3.3	513	US-11-128-049-5984	Sequence 5984, Ap
c	498	6	7	3.3	514	US-09-925-065A-288902	Sequence 288902,
	499	6	7	3.3	515	US-09-925-065A-243380	Sequence 243380,
	500	6	7	3.3	515	US-09-925-065A-476183	Sequence 476183,

ALIGNMENTS

RESULT 1
US-11-072-512-799
; Sequence 799, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO

APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHIKO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: Novel full length cDNA
FILE REFERENCE: 084335-0191
CURRENT APPLICATION NUMBER: US/11/072,512
CURRENT FILING DATE: 2005-03-07
PRIOR APPLICATION NUMBER: US 60/350,978
PRIOR FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: JP 2001-379298
PRIOR FILING DATE: 2001-11-05
NUMBER OF SEQ ID NOS: 4096
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 799
LENGTH: 1785
TYPE: DNA
ORGANISM: Homo sapiens
US-11-072-512-799

Alignment Scores:
Pred. No.:
Score: 6.63e-99   Length: 1785
Percent Similarity: 115.00   Matches: 211
Best Local Similarity: 99.1%   Conservative: 0
Query Match: 99.1%   Mismatches: 1
DB: 54.2%   Indels: 2
                          9   Gaps: 0

US-09-989-890-238 (1-212) x US-11-072-512-799 (1-1785)

Qy   1 SerProHisGlnAlaAlaAlaProValAspGlnThrProArgThrLeuAlaThrMetGly 20
Db 193 AGCCCCCACCAGGCCGCCGCCGCACCCGTAGACCAGACCCCAAGGACACCCTGGCCACCATGGGC 252

Qy   21 GlnArgAlaLeuProSerSerLeuAlaLeuLeuSerArgProLeuSerProProAla 40
Db 253 CAGAGAGAGCATTACCTTCATCTCTGCTCTGTCGCTGAGCCGGCCCTTGAGTCCCCACCTGCT 312

Qy   41 AlaCysSerGlyAspProGlyCysGlySerGlyAlaGlyLeuProSerAlaSerAlaAla 60
Db 313 GCCTGCTCTGGCGACCCTGGGTGTGGAGTGGTGCCGGCCTCGCCTTCTGCTTCCGCCGCT 372

Qy   61 AlaGlyIleAlaSerSerAlaValGluProValCysGlyAspAlaAlaProAlaCysLeu 80
Db 373 GCCGGGATTGCCTCCAGCCGTGGAGCCTGTGTGCGGGGATGCAGCCCCTGCCTGTCTA 432

Qy   81 LeuArgThrProLeuArgGlyLeuLeuLysProThrProArgSerThrMetGluCys 100
Db 433 CTGAGGACTCCACTGAGGGGACTGCTGAAGCCAACTGGGCCAAGGAGCACAATGGAGTGC 492

Qy 101 ProProAlaLeuIleValLeuHisProProAlaGlyGlyMetAlaSerGly-SerSerGlnPr 120
Db 493 CCCCCAGCCCTGATCGTGCACCCCCAGCCGCGGGAATGGCCAGCAG-CTCAAGTCAACC 551

Qy 120 oTrpAlaAlaSerAlaThrProMetLeuSerLeuSerAlaSerLeuCysIleProTh 140
Db 552 ATGGGCAGCAGAGCTTCAGCTACCCCGATGTTAAGCTCAAAGGCATCCCTGTATCCCTAC 611

Qy 140 rArgGlyProProGlnProLeuMetArgThrProAlaAlaArgSerHisTrpProThrProI 160
Db 612 CCGAGGGGCCACCTCCCCAGCCCCTGATGCGAACCCCTGCTCAAGGAGCCACTGGCCGAT 671

Qy 160 eProHisProCysAspThrAlaCysProAlaProLeuProValValLeuValAlaProAr 180
Db 672 CCCCCACCCATGCCACCTGGGACACAGCCCTGCCCAGCACCTTTGCCCAGTAGTCCTCGTGGCTCCGAG 731

Qy 180 gSerThrIleLeuSerMetSerArgThrTrpThrCysArgArgTrpAlaValAlaProCy 200
Db 732 GAGTACTATTCTTCCATGAGTCGGACCTGGACCTGCCGGAGGATGGGCAGTGGCTCCATG 791

Qy 200 sArgAlaGlyLysLeuMetCysSerSerArgSer 212
Db 792 TCGAGCCGAGAAATTGATGTGCTCATCTTCAAGAAGC 828

RESULT 2
US-11-124-368A-2250/c
; Sequence 2250, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2250
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-2250

Alignment Scores:
Pred. No.:
Score: 39.3   Length: 201
Percent Similarity: 9.00   Matches: 9
Best Local Similarity: 100.0%   Conservative: 0
Query Match: 100.0%   Mismatches: 0
DB: 4.2%   Indels: 0
                          12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-368A-2250 (1-201)

Qy   29 AlaLeuLeuSerArgProLeuSerPro 37
Db 172 GCTCTTCTTTCACGCCCACTTTCCCCA 146

RESULT 3
US-11-124-368A-2364/c
; Sequence 2364, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2364
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-2364

```
Alignment Scores:
Pred. No.:               39.3          Length:         201
Score:                   9.00          Matches:        9
Percent Similarity:      100.0%        Conservative:   0
Best Local Similarity:   100.0%        Mismatches:     0
Query Match:             4.2%          Indels:         0
DB:                      12            Gaps:           0

US-09-989-890-238 (1-212) x US-11-124-368A-2364 (1-201)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db   172 GCTCTTCTTTCACGGCCCACTTTCCCCA 146

RESULT 4
US-11-124-368A-2475/c
; Sequence 2475, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
;   APPLICANT: Michele Cargill
;   APPLICANT: James J. Devlin
;   APPLICANT: May Luke
;   TITLE OF INVENTION: Genetic Polymorphisms Associated with
;   TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
;   FILE REFERENCE: CL001524
;   CURRENT APPLICATION NUMBER: US/11/124,368A
;   CURRENT FILING DATE: 2005-05-09
;   PRIOR APPLICATION NUMBER: US 60/568,845
;   PRIOR FILING DATE: 2004-05-07
;   PRIOR APPLICATION NUMBER: US 60/625,936
;   PRIOR FILING DATE: 2004-11-09
;   NUMBER OF SEQ ID NOS: 21112
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2475
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-2475

Alignment Scores:
Pred. No.:               39.3          Length:         201
Score:                   9.00          Matches:        9
Percent Similarity:      100.0%        Conservative:   0
Best Local Similarity:   100.0%        Mismatches:     0
Query Match:             4.2%          Indels:         0
DB:                      12            Gaps:           0

US-09-989-890-238 (1-212) x US-11-124-368A-2475 (1-201)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db   172 GCTCTTCTTTCACGGCCCACTTTCCCCA 146

RESULT 5
US-11-124-368A-17847
; Sequence 17847, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
;   APPLICANT: Michele Cargill
;   APPLICANT: James J. Devlin
;   APPLICANT: May Luke
;   TITLE OF INVENTION: Genetic Polymorphisms Associated with
;   TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
;   FILE REFERENCE: CL001524
;   CURRENT APPLICATION NUMBER: US/11/124,368A
;   CURRENT FILING DATE: 2005-05-09
;   PRIOR APPLICATION NUMBER: US 60/568,845
;   PRIOR FILING DATE: 2004-05-07
;   PRIOR APPLICATION NUMBER: US 60/625,936
;   PRIOR FILING DATE: 2004-11-09
;   NUMBER OF SEQ ID NOS: 21112
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17847
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-17847

Alignment Scores:
Pred. No.:               39.3          Length:         201
Score:                   9.00          Matches:        9
Percent Similarity:      100.0%        Conservative:   0
Best Local Similarity:   100.0%        Mismatches:     0
Query Match:             4.2%          Indels:         0
DB:                      12            Gaps:           0

US-09-989-890-238 (1-212) x US-11-124-368A-17847 (1-201)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db    30 GCTCTTCTTTCACGGCCCACTTTCCCCA 56

RESULT 6
US-11-124-367A-3389/c
; Sequence 3389, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;   APPLICANT: Michele Cargill
;   APPLICANT: Hongjin Huang
;   TITLE OF INVENTION: Genetic Polymorphisms Associated with
;   TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
;   FILE REFERENCE: CL001519.ORD
;   CURRENT APPLICATION NUMBER: US/11/124,367A
;   CURRENT FILING DATE: 2005-05-09
;   PRIOR APPLICATION NUMBER: US 60/568,846
;   PRIOR FILING DATE: 2004-05-07
;   PRIOR APPLICATION NUMBER: US 60/582,609
;   PRIOR FILING DATE: 2004-06-25
;   PRIOR APPLICATION NUMBER: US 60/599,554
;   PRIOR FILING DATE: 2004-08-09
;   NUMBER OF SEQ ID NOS: 34460
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3389
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-367A-3389

Alignment Scores:
Pred. No.:               39.3          Length:         201
Score:                   9.00          Matches:        9
Percent Similarity:      100.0%        Conservative:   0
Best Local Similarity:   100.0%        Mismatches:     0
Query Match:             4.2%          Indels:         0
DB:                      12            Gaps:           0

US-09-989-890-238 (1-212) x US-11-124-367A-3389 (1-201)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db   172 GCTCTTCTTTCACGGCCCACTTTCCCCA 146

RESULT 7
US-11-124-367A-3491/c
; Sequence 3491, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;   APPLICANT: Michele Cargill
;   APPLICANT: Hongjin Huang
;   TITLE OF INVENTION: Genetic Polymorphisms Associated with
;   TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
;   FILE REFERENCE: CL001519.ORD
;   CURRENT APPLICATION NUMBER: US/11/124,367A
;   CURRENT FILING DATE: 2005-05-09
;   PRIOR APPLICATION NUMBER: US 60/568,846
;   PRIOR FILING DATE: 2004-05-07
```

; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3491
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-367A-3491

Alignment Scores:
Pred. No.:   Length: 201
Score: 39.3   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-367A-3491 (1-201)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   172 GCTCTTCTTTCACGCCCACTTCCCCA 146

RESULT 8
US-11-124-367A-3594/c
; Sequence 3594, Application US/1124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3594
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-367A-3594

Alignment Scores:
Pred. No.:   Length: 201
Score: 39.3   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-367A-3594 (1-201)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   172 GCTCTTCTTTCACGCCCACTTCCCCA 146

RESULT 9
US-11-124-367A-22366
; Sequence 22366, Application US/1124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang

; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22366
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-367A-22366

Alignment Scores:
Pred. No.:   Length: 201
Score: 39.3   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-367A-22366 (1-201)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   30 GCTCTTCTTTCACGCCCACTTCCCCA 56

RESULT 10
US-09-925-065A-468689/c
; Sequence 468689, Application US/09925065A
; Publication No. US2004018104811048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 468689
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-468689

Alignment Scores:
Pred. No.:   Length: 637
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-468689 (1-637)

Qy   25 ProSerSerLeuAlaLeuSerArg 33

Db   497 CCGAGCTCCTGCTCTACTATCCAGG 471

RESULT 11
US-09-925-065A-726141/c
; Sequence 726141, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 726141
;   LENGTH: 652
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-726141

Alignment Scores:
Pred. No.:   113   Length: 652
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-726141 (1-652)

Qy   28 LeuAlaLeuSerArgProLeuSer 36
            ||||||||||||||||||||||||
Db   576 TTAGCCCTTCTCTCTCGGCCTTTGTCT 550

RESULT 12
US-10-750-185-48878
; Sequence 48878, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;   APPLICANT: MMI GENOMICS, INC.
;   APPLICANT: DeNISE, Sue K.
;   APPLICANT: KERR, Richard
;   APPLICANT: ROSENFELD, David
;   APPLICANT: HOLM, Tom
;   APPLICANT: BATES, Stephen
;   APPLICANT: FANTIN, Dennis
;   TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;   FILE REFERENCE: MMI1100-2
;   CURRENT APPLICATION NUMBER: US/10/750,185
;   CURRENT FILING DATE: 2003-12-31
;   PRIOR APPLICATION NUMBER: US 60/437,482
;   PRIOR FILING DATE: 2002-12-31
;   NUMBER OF SEQ ID NOS: 64922
;   SOFTWARE: PatentIN version 3.1
;   SEQ ID NO 48878
;   LENGTH: 718
;   TYPE: DNA
;   ORGANISM: Bovine   1986680861922
US-10-750-185-48878

Alignment Scores:

Pred. No.:   123   Length: 718
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x US-10-750-185-48878 (1-718)

Qy   86 ArgGlyLeuLeuLysProThrGlyPro 94
            |||||||||||||||||||||||||
Db   128 AGGGGCGCTCCTGAAGCCTACTGGCCCG 154

RESULT 13
US-10-750-623-48878
; Sequence 48878, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
;   APPLICANT: MMI GENOMICS, INC.
;   APPLICANT: DeNISE, Sue K.
;   APPLICANT: KERR, Richard
;   APPLICANT: ROSENFELD, David
;   APPLICANT: HOLM, Tom
;   APPLICANT: BATES, Stephen
;   APPLICANT: FANTIN, Dennis
;   TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
;   FILE REFERENCE: MMI1100-1
;   CURRENT APPLICATION NUMBER: US/10/750,623
;   CURRENT FILING DATE: 2003-12-31
;   PRIOR APPLICATION NUMBER: US 60/437,482
;   PRIOR FILING DATE: 2002-12-31
;   NUMBER OF SEQ ID NOS: 64922
;   SOFTWARE: PatentIN version 3.1
;   SEQ ID NO 48878
;   LENGTH: 718
;   TYPE: DNA
;   ORGANISM: Bovine   1986680861922
US-10-750-623-48878

Alignment Scores:
Pred. No.:   123   Length: 718
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 8   Gaps: 0

US-09-989-890-238 (1-212) x US-10-750-623-48878 (1-718)

Qy   86 ArgGlyLeuLeuLysProThrGlyPro 94
            |||||||||||||||||||||||||
Db   128 AGGGGCGCTCCTGAAGCCTACTGGCCCG 154

RESULT 14
US-10-649-591-9/c
; Sequence 9, Application US/10649591
; Publication No. US20060035237A1
; GENERAL INFORMATION:
;   APPLICANT: Markowitz, Sanford D.
;   TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
;   TITLE OF INVENTION: CATEGORIZING PATIENTS
;   FILE REFERENCE: CWRU-P03-003
;   CURRENT APPLICATION NUMBER: US/10/649,591
;   CURRENT FILING DATE: 2003-08-26
;   PRIOR APPLICATION NUMBER: US 10/229,345
;   PRIOR FILING DATE: 2002-08-26
;   PRIOR APPLICATION NUMBER: US 10/274,177
;   PRIOR FILING DATE: 2002-10-18
;   NUMBER OF SEQ ID NOS: 22
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 9
;   LENGTH: 1209
;   TYPE: DNA

  ; ORGANISM: Homo sapiens
  ; FEATURE:
  ; NAME/KEY: misc_feature
  ; LOCATION: 1161
  ; OTHER INFORMATION: n = A,T,C or G
US-10-649-591-9

Alignment Scores:
Pred. No.:   196   Length: 1209
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 7   Gaps: 0

US-09-989-890-238 (1-212) x US-10-649-591-9 (1-1209)

Qy   59 AlaAlaAlaGlyIleAlaSerSerAla 67
         |||||||||||||||||||||||
Db   242 GCTGCTGCCGGGATCGCCTCCTCCGCG 216

RESULT 15
US-09-925-065A-691773
; Sequence 691773, Application US/09925065A
; Publication No. US2004018l048A1
; GENERAL INFORMATION:
  ; APPLICANT: Wang, David G.
  ; TITLE OF INVENTION: Identification and Mapping of Single
  ; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
  ; FILE REFERENCE: 108827.135
  ; CURRENT APPLICATION NUMBER: US/09/925,065A
  ; CURRENT FILING DATE: 2001-08-08
  ; PRIOR APPLICATION NUMBER: US 60/243,096
  ; PRIOR FILING DATE: 2000-10-24
  ; PRIOR APPLICATION NUMBER: US 60/252,147
  ; PRIOR FILING DATE: 2000-11-20
  ; PRIOR APPLICATION NUMBER: US 60/250,092
  ; PRIOR FILING DATE: 2000-11-30
  ; PRIOR APPLICATION NUMBER: US 60/261,766
  ; PRIOR FILING DATE: 2001-01-16
  ; PRIOR APPLICATION NUMBER: US 60/289,846
  ; PRIOR FILING DATE: 2001-05-09
  ; NUMBER OF SEQ ID NOS: 957086
  ; SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 691773
  ; LENGTH: 2520
  ; TYPE: DNA
  ; ORGANISM: Homo sapiens
  ; FEATURE:
  ; NAME/KEY: misc_feature
  ; LOCATION: 1550, 1551
  ; OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-691773

Alignment Scores:
Pred. No.:   378   Length: 2520
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-691773 (1-2520)

Qy   23 AlaLeuProSerSerLeuAlaLeuLeu 31
         ||||||||||||||||||||||||
Db   93 GCTCTTCCAAGTAGTCTTGCTCTGCTG 119

RESULT 16
US-10-750-185-53796
; Sequence 53796, Application US/10750185
; Publication No. US2005026603A1
; GENERAL INFORMATION:
  ; APPLICANT: MMI GENOMICS, INC.
  ; APPLICANT: DeNISE, Sue K.
  ; APPLICANT: KERR, Richard
  ; APPLICANT: ROSENFELD, David
  ; APPLICANT: HOLM, Tom
  ; APPLICANT: BATES, Stephen
  ; APPLICANT: FANTIN, Dennis
  ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
  ; FILE REFERENCE: MMI1100-2
  ; CURRENT APPLICATION NUMBER: US/10/750,185
  ; CURRENT FILING DATE: 2003-12-31
  ; PRIOR APPLICATION NUMBER: US 60/437,482
  ; PRIOR FILING DATE: 2002-12-31
  ; NUMBER OF SEQ ID NOS: 64922
  ; SOFTWARE: PatentIN version 3.1
SEQ ID NO 53796
  ; LENGTH: 2888
  ; TYPE: DNA
  ; ORGANISM: Bovine 1986680672526
US-10-750-185-53796

Alignment Scores:
Pred. No.:   427   Length: 2888
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 8
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-750-185-53796 (1-2888)

Qy   26 SerSerLeuAlaLeuSerArgPro 34
         ||||||||||||||||||
Db   846 TCTTCCTTAGCACTTCTATCACGACCA 872

RESULT 17
US-10-750-623-53796
; Sequence 53796, Application US/10750623
; Publication No. US2005028753lA1
; GENERAL INFORMATION:
  ; APPLICANT: MMI GENOMICS, INC.
  ; APPLICANT: DeNISE, Sue K.
  ; APPLICANT: KERR, Richard
  ; APPLICANT: ROSENFELD, David
  ; APPLICANT: HOLM, Tom
  ; APPLICANT: BATES, Stephen
  ; APPLICANT: FANTIN, Dennis
  ; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
  ; FILE REFERENCE: MMI1100-1
  ; CURRENT APPLICATION NUMBER: US/10/750,623
  ; CURRENT FILING DATE: 2003-12-31
  ; PRIOR APPLICATION NUMBER: US 60/437,482
  ; PRIOR FILING DATE: 2002-12-31
  ; NUMBER OF SEQ ID NOS: 64922
  ; SOFTWARE: PatentIN version 3.1
SEQ ID NO 53796
  ; LENGTH: 2888
  ; TYPE: DNA
  ; ORGANISM: Bovine 1986680672526
US-10-750-623-53796

Alignment Scores:
Pred. No.:   427   Length: 2888
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 8
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-10-750-623-53796 (1-2888)

Qy   26 SerSerLeuAlaLeuSerArgPro 34
         ||||||||||||||||||

Db   846 TCTTCCTTAGCACTTCTATCACGACCA 872

RESULT 18
US-11-124-368A-137/c
; Sequence 137, Application US/1112436BA
; Publication No. US20050287559A1
; GENERAL INFORMATION:
;   APPLICANT: Michele Cargill
;   APPLICANT: James J. Devlin
;   APPLICANT: May Luke
;   TITLE OF INVENTION: Genetic Polymorphisms Associated with
;   TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
;   FILE REFERENCE: CL001524
;   CURRENT APPLICATION NUMBER: US/11/124,368A
;   CURRENT FILING DATE: 2005-05-09
;   PRIOR APPLICATION NUMBER: US 60/568,845
;   PRIOR APPLICATION NUMBER: US 60/625,936
;   PRIOR FILING DATE: 2004-05-07
;   PRIOR FILING DATE: 2004-11-09
;   NUMBER OF SEQ ID NOS: 21112
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 137
;   LENGTH: 11598
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-124-368A-137

Alignment Scores:
Pred. No.:   1.48e+03   Length: 11598
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-368A-137 (1-11598)

Qy   29 AlaLeuSerArgProLeuSerPro 37
               ||||||||||||||||||||||||||
Db   4232 GCTCTTCTTTCACGCCCACTTTCCCCA 4206

RESULT 19
US-11-124-367A-173/c
; Sequence 173, Application US/1112436 7A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;   APPLICANT: Michele Cargill
;   APPLICANT: Hongjin Huang
;   TITLE OF INVENTION: Genetic Polymorphisms Associated with
;   TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
;   FILE REFERENCE: CL001519.ORD
;   CURRENT APPLICATION NUMBER: US/11/124,367A
;   CURRENT FILING DATE: 2005-05-09
;   PRIOR APPLICATION NUMBER: US 60/568,846
;   PRIOR APPLICATION NUMBER: US 60/582,609
;   PRIOR FILING DATE: 2004-05-07
;   PRIOR FILING DATE: 2004-06-25
;   PRIOR APPLICATION NUMBER: US 60/599,554
;   PRIOR FILING DATE: 2004-08-09
;   NUMBER OF SEQ ID NOS: 34460
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 173
;   LENGTH: 11598
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
US-11-124-367A-173

Alignment Scores:
Pred. No.:   1.48e+03   Length: 11598
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0

Db: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-367A-173 (1-11598)

Qy   29 AlaLeuSerArgProLeuSerPro 37
               ||||||||||||||||||||||||||
Db   4232 GCTCTTCTTTCACGCCCACTTTCCCCA 4206

RESULT 20
US-11-124-368A-136/c
; Sequence 136, Application US/1112436BA
; Publication No. US20050287559A1
; GENERAL INFORMATION:
;   APPLICANT: Michele Cargill
;   APPLICANT: James J. Devlin
;   APPLICANT: May Luke
;   TITLE OF INVENTION: Genetic Polymorphisms Associated with
;   TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
;   FILE REFERENCE: CL001524
;   CURRENT APPLICATION NUMBER: US/11/124,368A
;   CURRENT FILING DATE: 2005-05-09
;   PRIOR APPLICATION NUMBER: US 60/568,845
;   PRIOR APPLICATION NUMBER: US 60/625,936
;   PRIOR FILING DATE: 2004-05-07
;   PRIOR FILING DATE: 2004-11-09
;   NUMBER OF SEQ ID NOS: 21112
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 136
;   LENGTH: 11602
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-124-368A-136

Alignment Scores:
Pred. No.:   1.48e+03   Length: 11602
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-368A-136 (1-11602)

Qy   29 AlaLeuSerArgProLeuSerPro 37
               ||||||||||||||||||||||||||
Db   4236 GCTCTTCTTTCACGCCCACTTTCCCCA 4210

RESULT 21
US-11-124-367A-172/c
; Sequence 172, Application US/1112436 7A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;   APPLICANT: Michele Cargill
;   APPLICANT: Hongjin Huang
;   TITLE OF INVENTION: Genetic Polymorphisms Associated with
;   TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
;   FILE REFERENCE: CL001519.ORD
;   CURRENT APPLICATION NUMBER: US/11/124,367A
;   CURRENT FILING DATE: 2005-05-09
;   PRIOR APPLICATION NUMBER: US 60/568,846
;   PRIOR APPLICATION NUMBER: US 60/582,609
;   PRIOR FILING DATE: 2004-05-07
;   PRIOR FILING DATE: 2004-06-25
;   PRIOR APPLICATION NUMBER: US 60/599,554
;   PRIOR FILING DATE: 2004-08-09
;   NUMBER OF SEQ ID NOS: 34460
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 172
;   LENGTH: 11602
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
US-11-124-367A-172

Alignment Scores:
Pred. No.:   1.48e+03   Length: 11602
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-367A-172 (1-11602)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   4236 GCTCTTCTTTCACGCCCACTTTCCCCA 4210

RESULT 22
US-11-124-368A-135/c
; Sequence 135, Application US/1112436BA
; Publication No. US2005028755BA1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-05-07
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 135
;   LENGTH: 12678
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-124-368A-135

Alignment Scores:
Pred. No.:   1.6e+03   Length: 12678
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-368A-135 (1-12678)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   5312 GCTCTTCTTTCACGCCCACTTTCCCCA 5286

RESULT 23
US-11-124-367A-171/c
; Sequence 171, Application US/1112436 7A
; Publication No. US2006002470 0A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-08-09
; PRIOR APPLICATION NUMBER: US 60/599,554
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 171
;   LENGTH: 12678
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-124-367A-171

Alignment Scores:
Pred. No.:   1.6e+03   Length: 12678
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-367A-171 (1-12678)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   5312 GCTCTTCTTTCACGCCCACTTTCCCCA 5286

RESULT 24
US-11-124-368A-2920/c
; Sequence 2920, Application US/1112436BA
; Publication No. US2005028755BA1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-05-07
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2920
;   LENGTH: 41041
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-124-368A-2920

Alignment Scores:
Pred. No.:   4.59e+03   Length: 41041
Score: 9.00   Matches: 9
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 4.2%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-124-368A-2920 (1-41041)

Qy   29 AlaLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   5068 GCTCTTCTTTCACGCCCACTTTCCCCA 25042

RESULT 25
US-11-124-367A-5068/c
; Sequence 5068, Application US/1112436 7A
; Publication No. US2006002470 0A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07

```
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5068
;   LENGTH: 41041
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-124-367A-5068

Alignment Scores:
Pred. No.:              4.59e+03     Length:         41041
Score:                  9.00         Matches:        9
Percent Similarity:     100.0%       Conservative:   0
Best Local Similarity:  100.0%       Mismatches:     0
Query Match:            4.2%         Indels:         0
DB:                     12           Gaps:           0

US-09-989-890-238 (1-212) x US-11-124-367A-5068 (1-41041)

Qy    29 AlaLeuSerArgProLeuSerPro 37
         ||||||||||||||||||||||||
Db 25068 GCTCTTCTTTCACGCCCCACTTTCCCCA 25042

RESULT 26
US-10-330-773-261
; Sequence 261, Application US/10330773
; Publication No. US20060040262A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 261
;   LENGTH: 215248
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(215248)
;   OTHER INFORMATION: n = A,T,C or G
US-10-330-773-261

Alignment Scores:
Pred. No.:              2.02e+04     Length:         215248
Score:                  9.00         Matches:        9
Percent Similarity:     100.0%       Conservative:   0
Best Local Similarity:  100.0%       Mismatches:     0
Query Match:            4.2%         Indels:         0
DB:                     7            Gaps:           0

US-09-989-890-238 (1-212) x US-10-330-773-261 (1-215248)

Qy     26 SerSerLeuAlaLeuLeuSerArgPro 34
          |||||||||||||||||||||||||
Db 132221 TCCAGCTTGGCACTTCTGTCCAGGCCA 132247

RESULT 27
US-11-136-527-359961
; Sequence 359961, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
```

```
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 359961
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Probe
US-11-136-527-359961

Alignment Scores:
Pred. No.:              54.2         Length:         25
Score:                  8.00         Matches:        8
Percent Similarity:     100.0%       Conservative:   0
Best Local Similarity:  100.0%       Mismatches:     0
Query Match:            3.8%         Indels:         0
DB:                     12           Gaps:           0

US-09-989-890-238 (1-212) x US-11-136-527-359961 (1-25)

Qy   22 ArgAlaLeuProSerSerLeuAla 29
        |||||||||||||||||||||||||
Db    2 AGAGCCTGCCAAGTTCCTTGGCC 25

RESULT 28
US-11-103-122-44/c
; Sequence 44, Application US/11103122
; Publication No. US20050282190A1
; GENERAL INFORMATION:
; APPLICANT: Shi, Hua
; APPLICANT: Lis, John T.
; TITLE OF INVENTION: MODULAR DESIGN AND CONSTRUCTION OF NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, APTAMER-DERIVED NUCLEIC ACID CONSTRUCTS, RNA
; TITLE OF INVENTION: SCAFFOLDS, THEIR EXPRESSION, AND METHODS OF USE
; FILE REFERENCE: 19603/4491
; CURRENT APPLICATION NUMBER: US/11/103,122
; CURRENT FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: 60/560,895
; PRIOR FILING DATE: 2004-04-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
;   LENGTH: 198
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: Aptabody-B4/S1/101
US-11-103-122-44

Alignment Scores:
Pred. No.:              345          Length:         198
Score:                  8.00         Matches:        8
Percent Similarity:     100.0%       Conservative:   0
Best Local Similarity:  100.0%       Mismatches:     0
Query Match:            3.8%         Indels:         0
DB:                     12           Gaps:           0

US-09-989-890-238 (1-212) x US-11-103-122-44 (1-198)

Qy   36 SerProProAlaAlaCysSer 43
        |||||||||||||||||||||
Db   86 TCGCCGCCTCCGGCAGCTTGTAGT 63

RESULT 29
US-11-124-367A-24790
; Sequence 24790, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
```

```
;    APPLICANT: Hongjin Huang
;    TITLE OF INVENTION: Genetic Polymorphisms Associated with
;    TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
;    FILE REFERENCE: CL001519.ORD
;    CURRENT APPLICATION NUMBER: US/11/124,367A
;    CURRENT FILING DATE: 2005-05-09
;    PRIOR APPLICATION NUMBER: US 60/568,846
;    PRIOR FILING DATE: 2004-05-07
;    PRIOR APPLICATION NUMBER: US 60/582,609
;    PRIOR FILING DATE: 2004-06-25
;    PRIOR APPLICATION NUMBER: US 60/599,554
;    PRIOR FILING DATE: 2004-08-09
;    NUMBER OF SEQ ID NOS: 34460
;    SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 24790
;    LENGTH: 201
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-11-124-367A-24790

Alignment Scores:
Pred. No.:               350       Length:           201
Score:                  8.00       Matches:            8
Percent Similarity:   100.0%       Conservative:       0
Best Local Similarity:100.0%       Mismatches:         0
Query Match:            3.8%       Indels:             0
DB:                       12       Gaps:               0

US-09-989-890-238 (1-212) x US-11-124-367A-24790 (1-201)

Qy    33 ArgProLeuSerProProProAla 40
         ||||||||||||||||||||||||
Db    93 AGACCACTSTCCCCTCCACCTGCC 116


RESULT 30
US-11-124-367A-24791
; Sequence 24791, Application US/11124367A
; Publication No. US2006002470001A1
; GENERAL INFORMATION:
;    APPLICANT: Michele Cargill
;    APPLICANT: Hongjin Huang
;    TITLE OF INVENTION: Genetic Polymorphisms Associated with
;    TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
;    FILE REFERENCE: CL001519.ORD
;    CURRENT APPLICATION NUMBER: US/11/124,367A
;    CURRENT FILING DATE: 2005-05-09
;    PRIOR APPLICATION NUMBER: US 60/568,846
;    PRIOR FILING DATE: 2004-05-07
;    PRIOR APPLICATION NUMBER: US 60/582,609
;    PRIOR FILING DATE: 2004-06-25
;    PRIOR APPLICATION NUMBER: US 60/599,554
;    PRIOR FILING DATE: 2004-08-09
;    NUMBER OF SEQ ID NOS: 34460
;    SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 24791
;    LENGTH: 201
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-11-124-367A-24791

Alignment Scores:
Pred. No.:               350       Length:           201
Score:                  8.00       Matches:            8
Percent Similarity:   100.0%       Conservative:       0
Best Local Similarity:100.0%       Mismatches:         0
Query Match:            3.8%       Indels:             0
DB:                       12       Gaps:               0

US-09-989-890-238 (1-212) x US-11-124-367A-24791 (1-201)

Qy    33 ArgProLeuSerProProProAla 40
         ||||||||||||||||||||||||
Db   123 AGACCACTCTCCCCTCCACCTGCC 146


RESULT 31
US-11-124-367A-25234
; Sequence 25234, Application US/11124367A
; Publication No. US2006002470001A1
; GENERAL INFORMATION:
;    APPLICANT: Michele Cargill
;    APPLICANT: Hongjin Huang
;    TITLE OF INVENTION: Genetic Polymorphisms Associated with
;    TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
;    FILE REFERENCE: CL001519.ORD
;    CURRENT APPLICATION NUMBER: US/11/124,367A
;    CURRENT FILING DATE: 2005-05-09
;    PRIOR APPLICATION NUMBER: US 60/568,846
;    PRIOR FILING DATE: 2004-05-07
;    PRIOR APPLICATION NUMBER: US 60/582,609
;    PRIOR FILING DATE: 2004-06-25
;    PRIOR APPLICATION NUMBER: US 60/599,554
;    PRIOR FILING DATE: 2004-08-09
;    NUMBER OF SEQ ID NOS: 34460
;    SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 25234
;    LENGTH: 201
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-11-124-367A-25234

Alignment Scores:
Pred. No.:               350       Length:           201
Score:                  8.00       Matches:            8
Percent Similarity:   100.0%       Conservative:       0
Best Local Similarity:100.0%       Mismatches:         0
Query Match:            3.8%       Indels:             0
DB:                       12       Gaps:               0

US-09-989-890-238 (1-212) x US-11-124-367A-25234 (1-201)

Qy    33 ArgProLeuSerProProProAla 40
         ||||||||||||||||||||||||
Db   171 AGACCACTCTCCCCTCCACCTGCC 194


RESULT 32
US-09-925-065A-620441/c
; Sequence 620441, Application US/09925065A
; Publication No. US2004018104801A1
; GENERAL INFORMATION:
;    APPLICANT: Wang, David G.
;    TITLE OF INVENTION: Identification and Mapping of Single
;    TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;    FILE REFERENCE: 108827.135
;    CURRENT APPLICATION NUMBER: US/09/925,065A
;    CURRENT FILING DATE: 2001-08-08
;    PRIOR APPLICATION NUMBER: US 60/243,096
;    PRIOR FILING DATE: 2000-10-24
;    PRIOR APPLICATION NUMBER: US 60/252,147
;    PRIOR FILING DATE: 2000-11-20
;    PRIOR APPLICATION NUMBER: US 60/250,092
;    PRIOR FILING DATE: 2000-11-30
;    PRIOR APPLICATION NUMBER: US 60/261,766
;    PRIOR FILING DATE: 2001-01-16
;    PRIOR APPLICATION NUMBER: US 60/289,846
;    PRIOR FILING DATE: 2001-05-09
;    NUMBER OF SEQ ID NOS: 957086
;    SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 620441
;    LENGTH: 317
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: 200
;    OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-620441
```

Alignment Scores:
Pred. No.:
Score: 526
Percent Similarity: 8.00
Best Local Similarity: 100.0%
Query Match: 100.0%
DB: 3.8%
Length: 317
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 6

US-09-989-890-238 (1-212) x US-09-925-065A-620441 (1-317)

Qy   55 ProSerAlaSerAlaAlaGly 62
         |||||||||||||||||||||
Db   73 CCCTCAGCCTCAGCAGCAGGT 50

RESULT 33
US-09-925-065A-363884
; Sequence 363884, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 363884
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-363884

Alignment Scores:
Pred. No.:
Score: 614
Percent Similarity: 8.00
Best Local Similarity: 100.0%
Query Match: 100.0%
DB: 3.8%
Length: 377
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 6

US-09-989-890-238 (1-212) x US-09-925-065A-363884 (1-377)

Qy   34 ProLeuSerProProAlaAla 41
         |||||||||||||||||||||
Db   54 CCTCTCTCCTCCTCCTGCTGCT 77

RESULT 34
US-09-925-065A-363885
; Sequence 363885, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20

PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 363885
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-363885

Alignment Scores:
Pred. No.: 614
Score: 8.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 3.8%
DB: 6
Length: 377
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-363885 (1-377)

Qy   34 ProLeuSerProProAlaAla 41
         |||||||||||||||||||||
Db   54 CCTCTCTCCTCCTCCTGCTGCT 77

RESULT 35
US-09-925-065A-363886
; Sequence 363886, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 363886
;   LENGTH: 377
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-363886

Alignment Scores:
Pred. No.: 614
Score: 8.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 3.8%
DB: 6
Length: 377
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-363886 (1-377)

Qy   34 ProLeuSerProProAlaAla 41
         |||||||||||||||||||||
Db   54 CCTCTCTCCTCCTCCTGCTGCT 77

RESULT 36
US-11-043-752-949/c

; Sequence 949, Application US/11043752
; Publication No. US2006001416SA1
; GENERAL INFORMATION:
; APPLICANT: Hakonarson, Hakon
; APPLICANT: Gurney, Mark E.
; APPLICANT: Halapi, Eva
; TITLE OF INVENTION: METHODS OF DIAGNOSIS AND TREATMENT FOR
; TITLE OF INVENTION: ASTHMA AND OTHER RESPIRATORY DISEASES BASED ON HAPLOTYPE
; TITLE OF INVENTION: ASSOCIATION
; FILE REFERENCE: 2345.2044-003
; CURRENT APPLICATION NUMBER: US/11/043,752
; CURRENT FILING DATE: 2005-01-26
; PRIOR APPLICATION NUMBER: PCT/US04/022446
; PRIOR FILING DATE: 2004-07-14
; PRIOR APPLICATION NUMBER: 60/487,072
; PRIOR FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: 60/559,611
; PRIOR FILING DATE: 2004-04-05
; NUMBER OF SEQ ID NOS: 4326
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 949
; LENGTH: 395
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-043-752-949

Alignment Scores:
Pred. No.:   640
Score: 8.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 3.8%
DB: 12

US-09-989-890-238 (1-212) x US-11-043-752-949 (1-395)

Qy   28 LeuAlaLeuSerArgProLeu 35
               ||||||||||||||||||||||
Db   234 CTGGCTTTATTGAGCAGGCCCTTA 211

RESULT 37
US-11-000-688-608
; Sequence 608, Application US/11000688
; Publication No. US2005028754 4A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, Francois
; APPLICANT: HOULGATTE, Remi
; APPLICANT: BIRNBAUM, Daniel
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF COLON CANCER WITH DNA ARRAYS
; FILE REFERENCE: 1423-R-03
; CURRENT APPLICATION NUMBER: US/11/000,688
; CURRENT FILING DATE: 2004-12-01
; PRIOR APPLICATION NUMBER: US 60/525,987
; PRIOR FILING DATE: 2003-12-01
; NUMBER OF SEQ ID NOS: 1596
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 608
; LENGTH: 406
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequences:primer
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(406)
; OTHER INFORMATION: 5' terminal sequence from clone
; OTHER INFORMATION: image:283375. calcium channel, voltage-dependent,
; OTHER INFORMATION: alpha 1i subunit(CACNA1I) gene.
US-11-000-688-608

Alignment Scores:
Pred. No.:   656
Score: 8.00

Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-000-688-608 (1-406)

Qy   20 GlyGlnArgAlaLeuProSerSer 27
               ||||||||||||||||||||||||
Db   263 GGGCAAAGGGCCCTTCCTTCCTCC 286

RESULT 38
US-09-925-065A-844771/c
; Sequence 844771, Application US/09925065A
; Publication No. US2004018104 8A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 844771
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-844771

Alignment Scores:
Pred. No.:   659   Length: 408
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-844771 (1-408)

Qy   32 SerArgProLeuSerProProPro 39
               ||||||||||||||||||||||||
Db   308 TCCCGCCCACTCTCCCCACCTCCC 285

RESULT 39
US-10-623-155-97
; Sequence 97, Application US/10623155
; Publication No. US2005026116 6A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Peckham, David W.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C20
; CURRENT APPLICATION NUMBER: US/10/623,155
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 441
; TYPE: DNA

Length: 395
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 12, 308
;   OTHER INFORMATION: n = A,T,C or G
US-10-623-155-97

Alignment Scores:
Pred. No.:                        707          Length:          441
Score:                            8.00         Matches:         8
Percent Similarity:               100.0%       Conservative:    0
Best Local Similarity:            100.0%       Mismatches:      0
Query Match:                      3.8%         Indels:          0
DB:                               8            Gaps:            0

US-09-989-890-238 (1-212) x US-10-623-155-97 (1-441)

Qy      55 ProSerAlaSerAlaAlaAlaGly 62
           ||||||||||||||||||||||
Db      92 CCCTCAGCTTCAGCAGCAGCAGGT 115

RESULT 40
US-09-925-065A-489934/c
; Sequence 489934, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 489934
;   LENGTH: 473
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-489934

Alignment Scores:
Pred. No.:                        752          Length:          473
Score:                            8.00         Matches:         8
Percent Similarity:               100.0%       Conservative:    0
Best Local Similarity:            100.0%       Mismatches:      0
Query Match:                      3.8%         Indels:          0
DB:                               6            Gaps:            0

US-09-989-890-238 (1-212) x US-09-925-065A-489934 (1-473)

Qy      26 SerSerLeuAlaLeuLeuSerArg 33
           ||||||||||||||||||||||
Db     268 AGCTCTTTGGCTCTGCTCAGCAGG 245

RESULT 41
US-10-667-295-223/c
; Sequence 223, Application US/10667295
; Publication No. US2005025725A1
; GENERAL INFORMATION:
;   APPLICANT: Mascia, Peter
;   TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
;   FILE REFERENCE: 11696-047001
;   CURRENT APPLICATION NUMBER: US/10/667,295
```

```
;   CURRENT FILING DATE: 2003-09-17
;   PRIOR APPLICATION NUMBER: US 60/411,823
;   PRIOR FILING DATE: 2002-09-17
;   NUMBER OF SEQ ID NOS: 263
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 223
;   LENGTH: 475
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(475)
;   OTHER INFORMATION: Ceres Seq. ID no. 13633589
US-10-667-295-223

Alignment Scores:
Pred. No.:                        755          Length:          475
Score:                            8.00         Matches:         8
Percent Similarity:               100.0%       Conservative:    0
Best Local Similarity:            100.0%       Mismatches:      0
Query Match:                      3.8%         Indels:          0
DB:                               8            Gaps:            0

US-09-989-890-238 (1-212) x US-10-667-295-223 (1-475)

Qy      33 SerSerLeuAlaLeuSerArg 33
           |||||||||||||||||||||
Db     403 TCCTCCCTAGCCCTCCTTTCAAGA 380

RESULT 42
US-10-667-295-15/c
; Sequence 15, Application US/10667295
; Publication No. US2005025725A1
; GENERAL INFORMATION:
;   APPLICANT: Mascia, Peter
;   TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
;   FILE REFERENCE: 11696-047001
;   CURRENT APPLICATION NUMBER: US/10/667,295
;   CURRENT FILING DATE: 2003-09-17
;   PRIOR APPLICATION NUMBER: US 60/411,823
;   PRIOR FILING DATE: 2002-09-17
;   NUMBER OF SEQ ID NOS: 263
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 15
;   LENGTH: 497
;   TYPE: DNA
;   ORGANISM: Brassica napus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(497)
;   OTHER INFORMATION: Ceres Seq. ID no. 12474809
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(497)
;   OTHER INFORMATION: n = A,T,C or G
US-10-667-295-15

Alignment Scores:
Pred. No.:                        786          Length:          497
Score:                            8.00         Matches:         8
Percent Similarity:               100.0%       Conservative:    0
Best Local Similarity:            100.0%       Mismatches:      0
Query Match:                      3.8%         Indels:          0
DB:                               8            Gaps:            0

US-09-989-890-238 (1-212) x US-10-667-295-15 (1-497)

Qy      26 SerSerLeuAlaLeuLeuSerArg 33
           ||||||||||||||||||||||
Db     468 TCCTCCCTAGCCCTCCTTTCAAGA 445

RESULT 43
US-09-925-065A-616675
```

```
; Sequence 616675, Application US/09925065A
; Publication No. US2004018I1048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 616675
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-616675

Alignment Scores:
Pred. No.:                815         Length:        517
Score:                    8.00        Matches:       8
Percent Similarity:       100.0%      Conservative:  0
Best Local Similarity:    100.0%      Mismatches:    0
Query Match:              3.8%        Indels:        0
DB:                       6           Gaps:          0

US-09-989-890-238 (1-212) x US-09-925-065A-616675 (1-517)

Qy   58 SerAlaAlaAlaGlyIleAlaSer 65
         ||||||||||||||||||||||||
Db  202 TCTGCAGCTGCTGGAATAGCTTCG 225

RESULT 44
US-09-925-065A-616676
; Sequence 616676, Application US/09925065A
; Publication No. US2004018I1048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 616676
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-616676

Alignment Scores:
Pred. No.:                815         Length:        517
Score:                    8.00        Matches:       8
Percent Similarity:       100.0%      Conservative:  0
Best Local Similarity:    100.0%      Mismatches:    0
Query Match:              3.8%        Indels:        0
DB:                       6           Gaps:          0

US-09-989-890-238 (1-212) x US-09-925-065A-616676 (1-517)

Qy   58 SerAlaAlaAlaGlyIleAlaSer 65
         ||||||||||||||||||||||||
Db  202 TCTGCAGCTGCTGGAATAGCTTCG 225

RESULT 45
US-09-925-065A-418469/c
; Sequence 418469, Application US/09925065A
; Publication No. US2004018I1048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 418469
;   LENGTH: 518
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-418469

Alignment Scores:
Pred. No.:                816         Length:        518
Score:                    8.00        Matches:       8
Percent Similarity:       100.0%      Conservative:  0
Best Local Similarity:    100.0%      Mismatches:    0
Query Match:              3.8%        Indels:        0
DB:                       6           Gaps:          0

US-09-989-890-238 (1-212) x US-09-925-065A-418469 (1-518)

Qy   23 AlaLeuProSerSerLeuAlaLeu 30
         ||||||||||||||||||||||||
Db  260 GCTCTGCCCTCTAGCCTTGCTCTC 237

RESULT 46
US-10-467-657-2511/c
; Sequence 2511, Application US/10467657
; Publication No. US2005026058IA1
; GENERAL INFORMATION:
;   APPLICANT: CHIRON SpA
;   APPLICANT: FONTANA Maria Rita
;   APPLICANT: PIZZA Mariagrazia
;   APPLICANT: MASIGNANI Vega
;   APPLICANT: MONACI Elisabetta
;   TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;   FILE REFERENCE:
;   CURRENT APPLICATION NUMBER: US/10/467,657
;   CURRENT FILING DATE: 2003-08-11
;   PRIOR APPLICATION NUMBER: GB-0103424.8
;   PRIOR FILING DATE: 2001-02-12
;   NUMBER OF SEQ ID NOS: 9218
;   SOFTWARE: SeqWin99, version 1.04
;   SEQ ID NO 2511
```

```
;   LENGTH: 525
;   TYPE: DNA
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2511

Alignment Scores:
Pred. No.:                    826      Length:           525
Score:                        8.00     Matches:          8
Percent Similarity:           100.0%   Conservative:     0
Best Local Similarity:        100.0%   Mismatches:       0
Query Match:                  3.8%     Indels:           0
DB:                                    Gaps:             0

US-09-989-890-238 (1-212) x US-10-467-657-2511 (1-525)

Qy    55 ProSerAlaSerAlaAlaAlaGly 62
            ||||||||||||||||||||||
Db    62 CCTTCCGCCAGTGCGGCGGCCGG 39

RESULT 47
US-11-021-492-486/c
; Sequence 486, Application US/11021492
; Publication No. US20060031947A1
; GENERAL INFORMATION:
;   APPLICANT: Abuin, Alejandro
;   APPLICANT: Zambrowicz, Brian
;   APPLICANT: Sands, Arthur T.
;   TITLE OF INVENTION: Novel Mutated Mammalian Cells and
;   TITLE OF INVENTION: Animals
;   FILE REFERENCE: LEX-0368-USA
;   CURRENT APPLICATION NUMBER: US/11/021,492
;   CURRENT FILING DATE: 2004-12-23
;   PRIOR APPLICATION NUMBER: US 60/307,670
;   PRIOR FILING DATE: 2001-07-25
;   NUMBER OF SEQ ID NOS: 698
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 486
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 469, 478, 481, 514, 522, 526
;   OTHER INFORMATION: n = A,T,C or G
US-11-021-492-486

Alignment Scores:
Pred. No.:                    850      Length:           542
Score:                        8.00     Matches:          8
Percent Similarity:           100.0%   Conservative:     0
Best Local Similarity:        100.0%   Mismatches:       0
Query Match:                  3.8%     Indels:           0
DB:                                    Gaps:             9

US-09-989-890-238 (1-212) x US-11-021-492-486 (1-542)

Qy    44 GlyAspProGlyCysGlySerGly 51
            ||||||||||||||||||
Db   425 GGTGACCCCGGCTGTGGCTCCGGG 402

RESULT 48
US-09-925-065A-151898/c
; Sequence 151898, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
```

```
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 151898
;   LENGTH: 547
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-151898

Alignment Scores:
Pred. No.:                    857      Length:           547
Score:                        8.00     Matches:          8
Percent Similarity:           100.0%   Conservative:     0
Best Local Similarity:        100.0%   Mismatches:       0
Query Match:                  3.8%     Indels:           0
DB:                                    Gaps:             6

US-09-989-890-238 (1-212) x US-09-925-065A-151898 (1-547)

Qy    24 LeuProSerLeuAlaLeuLeu 31
            |||||||||||||||||||
Db    41 TTGCCTAGCCAGCCTAGCACTTTTG 18

RESULT 49
US-09-925-065A-748081/c
; Sequence 748081, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 748081
;   LENGTH: 552
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-748081

Alignment Scores:
Pred. No.:                    864      Length:           552
Score:                        8.00     Matches:          8
Percent Similarity:           100.0%   Conservative:     0
Best Local Similarity:        100.0%   Mismatches:       0
Query Match:                  3.8%     Indels:           0
DB:                                    Gaps:             6

US-09-989-890-238 (1-212) x US-09-925-065A-748081 (1-552)

Qy    24 LeuProSerLeuAlaLeuLeu 31
            |||||||||||||||||||
Db   183 TTGCCTAGCCAGCCTAGCACTTTTG 160
```

RESULT 50
US-09-925-065A-808255/c
; Sequence 808255, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 808255
;   LENGTH: 559
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-808255

Alignment Scores:
Pred. No.:   874   Length: 559
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-808255 (1-559)

Qy   54 LeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||
Db   317 CTCCCCAGTGCTTCTGCAGCCGCC 294

RESULT 51
US-09-925-065A-808256/c
; Sequence 808256, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 808256
;   LENGTH: 559
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-808256

Alignment Scores:

RESULT 52
US-09-925-065A-808257/c
; Sequence 808257, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 808257
;   LENGTH: 559
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-808257

Alignment Scores:
Pred. No.:   874   Length: 559
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-808256 (1-559)

Qy   54 LeuProSerAlaSerAlaAlaAla 61
           ||||||||||||||||||||||||
Db   317 CTCCCCAGTGCTTCTGCAGCCGCC 294

RESULT 53
US-09-925-065A-946604
; Sequence 946604, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30

```
;     PRIOR APPLICATION NUMBER: US 60/261,766
;     PRIOR FILING DATE: 2001-01-16
;     PRIOR APPLICATION NUMBER: US 60/289,846
;     PRIOR FILING DATE: 2001-05-09
;     NUMBER OF SEQ ID NOS: 957086
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 946604
;     LENGTH: 565
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-925-065A-946604

Alignment Scores:
Pred. No.:                882          Length:              565
Score:                   8.00         Matches:               8
Percent Similarity:     100.0%        Conservative:          0
Best Local Similarity:  100.0%        Mismatches:            0
Query Match:             3.8%         Indels:                0
DB:                        6          Gaps:                  0

US-09-989-890-238 (1-212) x US-09-925-065A-946604 (1-565)

Qy    54 LeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||
Db   490 CTCCCCTCTGCAAGTGCCGCAGCT 513

RESULT 54
US-09-925-065A-693639
; Sequence 693639, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;     APPLICANT: Wang, David G.
;     TITLE OF INVENTION: Identification and Mapping of Single
;     TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;     FILE REFERENCE: 108827.135
;     CURRENT APPLICATION NUMBER: US/09/925,065A
;     CURRENT FILING DATE: 2001-08-08
;     PRIOR APPLICATION NUMBER: US 60/243,096
;     PRIOR FILING DATE: 2000-10-24
;     PRIOR APPLICATION NUMBER: US 60/252,147
;     PRIOR FILING DATE: 2000-11-20
;     PRIOR APPLICATION NUMBER: US 60/261,766
;     PRIOR FILING DATE: 2001-01-16
;     PRIOR APPLICATION NUMBER: US 60/289,846
;     PRIOR FILING DATE: 2001-05-09
;     NUMBER OF SEQ ID NOS: 957086
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 693639
;     LENGTH: 570
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-925-065A-693639

Alignment Scores:
Pred. No.:                889          Length:              570
Score:                   8.00         Matches:               8
Percent Similarity:     100.0%        Conservative:          0
Best Local Similarity:  100.0%        Mismatches:            0
Query Match:             3.8%         Indels:                0
DB:                        6          Gaps:                  0

US-09-989-890-238 (1-212) x US-09-925-065A-693639 (1-570)

Qy    28 LeuAlaLeuLeuSerArgProLeu 35
         ||||||||||||||||||||||||
Db   473 CTTGCTTTGTTGTCGAGGCCGCTC 496

RESULT 55
US-09-925-065A-950741
; Sequence 950741, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;     APPLICANT: Wang, David G.
;     TITLE OF INVENTION: Identification and Mapping of Single
;     TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;     FILE REFERENCE: 108827.135
;     CURRENT APPLICATION NUMBER: US/09/925,065A
;     CURRENT FILING DATE: 2001-08-08
;     PRIOR APPLICATION NUMBER: US 60/243,096
;     PRIOR FILING DATE: 2000-10-24
;     PRIOR APPLICATION NUMBER: US 60/252,147
;     PRIOR FILING DATE: 2000-11-20
;     PRIOR APPLICATION NUMBER: US 60/261,766
;     PRIOR FILING DATE: 2001-01-16
;     PRIOR APPLICATION NUMBER: US 60/289,846
;     PRIOR FILING DATE: 2001-05-09
;     NUMBER OF SEQ ID NOS: 957086
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 950741
;     LENGTH: 570
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-925-065A-950741

Alignment Scores:
Pred. No.:                889          Length:              570
Score:                   8.00         Matches:               8
Percent Similarity:     100.0%        Conservative:          0
Best Local Similarity:  100.0%        Mismatches:            0
Query Match:             3.8%         Indels:                0
DB:                        6          Gaps:                  0

US-09-989-890-238 (1-212) x US-09-925-065A-950741 (1-570)

Qy    54 LeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||
Db   495 CTCCCCTCTGCAAGTGCCGCAGCT 518

RESULT 56
US-09-925-065A-733926/c
; Sequence 733926, Application US2004018104A1
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;     APPLICANT: Wang, David G.
;     TITLE OF INVENTION: Identification and Mapping of Single
;     TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;     FILE REFERENCE: 108827.135
;     CURRENT APPLICATION NUMBER: US/09/925,065A
;     CURRENT FILING DATE: 2001-08-08
;     PRIOR APPLICATION NUMBER: US 60/243,096
;     PRIOR FILING DATE: 2000-10-24
;     PRIOR APPLICATION NUMBER: US 60/252,147
;     PRIOR FILING DATE: 2000-11-20
;     PRIOR APPLICATION NUMBER: US 60/250,092
;     PRIOR FILING DATE: 2000-11-30
;     PRIOR APPLICATION NUMBER: US 60/261,766
;     PRIOR FILING DATE: 2001-01-16
;     PRIOR APPLICATION NUMBER: US 60/289,846
;     PRIOR FILING DATE: 2001-05-09
;     NUMBER OF SEQ ID NOS: 957086
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 733926
;     LENGTH: 571
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-925-065A-733926

Alignment Scores:
Pred. No.:                890          Length:              571
Score:                   8.00         Matches:               8
Percent Similarity:     100.0%        Conservative:          0
Best Local Similarity:  100.0%        Mismatches:            0
```

Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-733926 (1-571)

Qy   28 LeuAlaLeuLeuSerArgProLeu 35
         ||||||||||||||||||||||||
Db   131 CTTGCTTTGTTGTCGAGGCCGCTC 108

RESULT 57
US-09-925-065A-854936/c
; Sequence 854936, Application US/09925065A
; Publication No. US2004018l048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 854936
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-854936

Alignment Scores:
Pred. No.:   897   Length: 576
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-854936 (1-576)

Qy   27 SerLeuAlaLeuSerArgPro 34
         |||||||||||||||||||||
Db   484 AGTCTCGCTCTGTTGTCCAGGCCG 461

RESULT 58
US-09-925-065A-53760/c
; Sequence 53760, Application US/09925065A
; Publication No. US2004018l048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 53760
; LENGTH: 581
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-53760

Alignment Scores:
Pred. No.:   904   Length: 581
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-53760 (1-581)

Qy   54 LeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||
Db   157 CTCCCCTCTGCCTCCGGCTGCTGCA 134

RESULT 59
US-09-925-065A-53761/c
; Sequence 53761, Application US/09925065A
; Publication No. US2004018l048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 53761
; LENGTH: 581
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-53761

Alignment Scores:
Pred. No.:   904   Length: 581
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-53761 (1-581)

Qy   54 LeuProSerAlaSerAlaAlaAla 61
         ||||||||||||||||||||||||
Db   157 CTCCCCTCTGCCTCCGGCTGCTGCA 134

RESULT 60
US-09-925-065A-53762/c
; Sequence 53762, Application US/09925065A
; Publication No. US2004018l048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome

```
    ;   FILE REFERENCE: 108827.135
    ;   CURRENT APPLICATION NUMBER: US/09/925,065A
    ;   CURRENT FILING DATE: 2001-08-08
    ;   PRIOR APPLICATION NUMBER: US 60/243,096
    ;   PRIOR FILING DATE: 2000-10-24
    ;   PRIOR APPLICATION NUMBER: US 60/252,147
    ;   PRIOR FILING DATE: 2000-11-20
    ;   PRIOR APPLICATION NUMBER: US 60/250,092
    ;   PRIOR FILING DATE: 2000-11-30
    ;   PRIOR APPLICATION NUMBER: US 60/261,766
    ;   PRIOR FILING DATE: 2001-01-16
    ;   PRIOR APPLICATION NUMBER: US 60/289,846
    ;   PRIOR FILING DATE: 2001-05-09
    ;   NUMBER OF SEQ ID NOS: 957086
    ;   SOFTWARE: FastSEQ for Windows Version 4.0
    ;   SEQ ID NO 53762
    ;       LENGTH: 581
    ;       TYPE: DNA
    ;       ORGANISM: Homo sapiens
    US-09-925-065A-53762

    Alignment Scores:
    Pred. No.:                  904         Length:            581
    Score:                     8.00         Matches:             8
    Percent Similarity:       100.0%        Conservative:        0
    Best Local Similarity:    100.0%        Mismatches:          0
    Query Match:               3.8%         Indels:              0
    DB:                                     Gaps:                0

    US-09-989-890-238 (1-212) x US-09-925-065A-53762 (1-581)

    Qy      54  LeuProSerAlaSerAlaAlaAla 61
                ||||||||||||||||||||||||
    Db     157  CTCCCCTCTGCCTCGGCTGCTGCA 134


    RESULT 61
    US-09-925-065A-295614/c
    ; Sequence 295614, Application US/09925065A
    ; Publication No. US20040181048A1
    ; GENERAL INFORMATION:
    ;   APPLICANT: Wang, David G.
    ;   TITLE OF INVENTION: Identification and Mapping of Single
    ;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
    ;   FILE REFERENCE: 108827.135
    ;   CURRENT APPLICATION NUMBER: US/09/925,065A
    ;   CURRENT FILING DATE: 2001-08-08
    ;   PRIOR APPLICATION NUMBER: US 60/243,096
    ;   PRIOR FILING DATE: 2000-10-24
    ;   PRIOR APPLICATION NUMBER: US 60/252,147
    ;   PRIOR FILING DATE: 2000-11-20
    ;   PRIOR APPLICATION NUMBER: US 60/250,092
    ;   PRIOR FILING DATE: 2000-11-30
    ;   PRIOR APPLICATION NUMBER: US 60/261,766
    ;   PRIOR FILING DATE: 2001-01-16
    ;   PRIOR APPLICATION NUMBER: US 60/289,846
    ;   PRIOR FILING DATE: 2001-05-09
    ;   NUMBER OF SEQ ID NOS: 957086
    ;   SOFTWARE: FastSEQ for Windows Version 4.0
    ;   SEQ ID NO 295614
    ;       LENGTH: 586
    ;       TYPE: DNA
    ;       ORGANISM: Homo sapiens
    US-09-925-065A-295614

    Alignment Scores:
    Pred. No.:                  911         Length:            586
    Score:                     8.00         Matches:             8
    Percent Similarity:       100.0%        Conservative:        0
    Best Local Similarity:    100.0%        Mismatches:          0
    Query Match:               3.8%         Indels:              0
    DB:                                     Gaps:                0

    .   US-09-989-890-238 (1-212) x US-09-925-065A-295614 (1-586)
```

```
    Qy      53  GlyLeuProSerAlaSerAlaAla 60
                ||||||||||||||||||||||||
    Db     244  GGGCTCCCATCTGCCTCTGCTGCT 221


    RESULT 62
    US-09-925-065A-630739/c
    ; Sequence 630739, Application US/09925065A
    ; Publication No. US20040181048A1
    ; GENERAL INFORMATION:
    ;   APPLICANT: Wang, David G.
    ;   TITLE OF INVENTION: Identification and Mapping of Single
    ;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
    ;   FILE REFERENCE: 108827.135
    ;   CURRENT APPLICATION NUMBER: US/09/925,065A
    ;   CURRENT FILING DATE: 2001-08-08
    ;   PRIOR APPLICATION NUMBER: US 60/243,096
    ;   PRIOR FILING DATE: 2000-10-24
    ;   PRIOR APPLICATION NUMBER: US 60/252,147
    ;   PRIOR FILING DATE: 2000-11-20
    ;   PRIOR APPLICATION NUMBER: US 60/250,092
    ;   PRIOR FILING DATE: 2000-11-30
    ;   PRIOR APPLICATION NUMBER: US 60/261,766
    ;   PRIOR FILING DATE: 2001-01-16
    ;   PRIOR APPLICATION NUMBER: US 60/289,846
    ;   PRIOR FILING DATE: 2001-05-09
    ;   NUMBER OF SEQ ID NOS: 957086
    ;   SOFTWARE: FastSEQ for Windows Version 4.0
    ;   SEQ ID NO 630739
    ;       LENGTH: 593
    ;       TYPE: DNA
    ;       ORGANISM: Homo sapiens
    US-09-925-065A-630739

    Alignment Scores:
    Pred. No.:                  921         Length:            593
    Score:                     8.00         Matches:             8
    Percent Similarity:       100.0%        Conservative:        0
    Best Local Similarity:    100.0%        Mismatches:          0
    Query Match:               3.8%         Indels:              0
    DB:                                     Gaps:                0

    US-09-989-890-238 (1-212) x US-09-925-065A-630739 (1-593)

    Qy      25  ProSerSerLeuAlaLeuLeuSer 32
                ||||||||||||||||||||||||
    Db     445  CCATCCAGCTTGGCTCTGCTGTCG 422


    RESULT 63
    US-09-925-065A-295459
    ; Sequence 295459, Application US/09925065A
    ; Publication No. US20040181048A1
    ; GENERAL INFORMATION:
    ;   APPLICANT: Wang, David G.
    ;   TITLE OF INVENTION: Identification and Mapping of Single
    ;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
    ;   FILE REFERENCE: 108827.135
    ;   CURRENT APPLICATION NUMBER: US/09/925,065A
    ;   CURRENT FILING DATE: 2001-08-08
    ;   PRIOR APPLICATION NUMBER: US 60/243,096
    ;   PRIOR FILING DATE: 2000-10-24
    ;   PRIOR APPLICATION NUMBER: US 60/252,147
    ;   PRIOR FILING DATE: 2000-11-20
    ;   PRIOR APPLICATION NUMBER: US 60/250,092
    ;   PRIOR FILING DATE: 2000-11-30
    ;   PRIOR APPLICATION NUMBER: US 60/261,766
    ;   PRIOR FILING DATE: 2001-01-16
    ;   PRIOR APPLICATION NUMBER: US 60/289,846
    ;   PRIOR FILING DATE: 2001-05-09
    ;   NUMBER OF SEQ ID NOS: 957086
    ;   SOFTWARE: FastSEQ for Windows Version 4.0
    ;   SEQ ID NO 295459
    ;       LENGTH: 595
```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(595)
; OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-295459

Alignment Scores:
Pred. No.:   595
Score: 924   8
Percent Similarity: 8.00   0
Best Local Similarity: 100.0%   0
Query Match: 100.0%   0
DB: 3.8%
                             6

US-09-989-890-238 (1-212) x US-09-925-065A-295459 (1-595)

Qy   30 LeuLeuSerArgProLeuSerPro 37
           ||||||||||||||||||||||||
Db   507 TTGCTGTCTAGACCCCTTAGTCCT 530

RESULT 64
US-09-925-065A-641425
; Sequence 641425, Application US/09925065A
; Publication No. US2004018I1048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 641425
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 380
; OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-641425

Alignment Scores:
Pred. No.:   595
Score: 924   8
Percent Similarity: 8.00   0
Best Local Similarity: 100.0%   0
Query Match: 100.0%   0
DB: 3.8%
                             6

US-09-989-890-238 (1-212) x US-09-925-065A-641425 (1-595)

Qy   25 ProSerSerLeuAlaLeuLeuSer 32
           ||||||||||||||||||||||||
Db   152 CCATCCAGCTTGGCTCTGCTGTCG 175

RESULT 65
US-09-925-065A-301227
; Sequence 301227, Application US/09925065A
; Publication No. US2004018I1048A1

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 301227
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-301227

Alignment Scores:
Pred. No.:   599
Score: 929   8
Percent Similarity: 8.00   0
Best Local Similarity: 100.0%   0
Query Match: 100.0%   0
DB: 3.8%
                             6

US-09-989-890-238 (1-212) x US-09-925-065A-301227 (1-599)

Qy   170 AlaProLeuProValValLeuVal 177
           ||||||||||||||||||||||||
Db   560 GCTCCATTGCCAGTGGTTCTTGTC 583

RESULT 66
US-09-925-065A-301228
; Sequence 301228, Application US/09925065A
; Publication No. US2004018I1048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 301228
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-301228

Alignment Scores:
Pred. No.:   599
Score: 929   8
Percent Similarity: 8.00   0
Best Local Similarity: 100.0%

```
Percent Similarity:     100.0%     Conservative:     0
Best Local Similarity:  100.0%     Mismatches:       0
Query Match:            3.8%       Indels:           0
DB:                     12         Gaps:             0

US-09-989-890-238 (1-212) x US-11-136-527-8133 (1-600)

Qy      22 ArgAlaLeuProSerSerLeuAla 29
            ||||||||||||||||||||||||
Db     397 AGAGCCCTGCCAAGTTCCTTGGCC 420

RESULT 69
US-09-925-065A-774077/c
; Sequence 774077, Application US/09925065A
; Publication No. US2004018104sA1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE:  2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE:  2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE:  2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE:  2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE:  2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE:  2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 774077
;   LENGTH: 602
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-774077

Alignment Scores:
Pred. No.:                         Length:           602
Score:                  8.00       Matches:          8
Percent Similarity:     100.0%     Conservative:     0
Best Local Similarity:  100.0%     Mismatches:       0
Query Match:            3.8%       Indels:           0
DB:                     6          Gaps:             0

US-09-989-890-238 (1-212) x US-09-925-065A-774077 (1-602)

Qy      25 ProSerSerAlaAlaLeuLeuSer 32
            ||||||||||||||||||||||||
Db     198 CCCAGCTCTTAGCTCTCTTGAGT 175

RESULT 70
US-09-925-065A-837176/c
; Sequence 837176, Application US/09925065A
; Publication No. US2004018104sA1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE:  2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE:  2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE:  2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE:  2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE:  2001-01-16
```

```
Query Match:            3.8%       Indels:           0
DB:                     6          Gaps:             0

US-09-989-890-238 (1-212) x US-09-925-065A-301228 (1-599)

Qy     170 AlaProLeuProValValLeuVal 177
            ||||||||||||||||||||||||
Db     560 GCTCCATTGCCAGTGGTTCTTGTC 583

RESULT 67
US-10-623-155-98
; Sequence 98, Application US/10623155
; Publication No. US20050261166A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, Tongtong
;   APPLICANT: Peckham, David W.
;   APPLICANT: Retter, Marc W.
;   APPLICANT: Fanger, Gary R.
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;   TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;   FILE REFERENCE: 210121.455C20
;   CURRENT APPLICATION NUMBER: US/10/623,155
;   CURRENT FILING DATE:  2003-07-17
;   NUMBER OF SEQ ID NOS: 560
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 98
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 295_349, 489, 496, 583
;   OTHER INFORMATION: n = A,T,C or G
US-10-623-155-98

Alignment Scores:
Pred. No.:                         Length:           600
Score:                  931        Matches:          8
Percent Similarity:     8.00       Conservative:     0
Best Local Similarity:  100.0%     Mismatches:       0
Query Match:            100.0%     Indels:           0
DB:                     3.8%       Gaps:             0
                        8

US-09-989-890-238 (1-212) x US-10-623-155-98 (1-600)

Qy      55 ProSerAlaSerAlaAlaAlaGly 62
            ||||||||||||||||||||||||
Db      79 CCCTCAGCTTCAGCAGCAGCAGGT 102

RESULT 68
US-11-136-527-8133
; Sequence 8133, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
;   APPLICANT: Wyeth
;   APPLICANT: Mounts, William M
;   TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
;   FILE REFERENCE: 031896-041000 (AM101086)
;   CURRENT APPLICATION NUMBER: US/11/136,527
;   CURRENT FILING DATE:  2005-05-25
;   PRIOR APPLICATION NUMBER: US 60/574,294
;   PRIOR FILING DATE:  2005-05-26
;   NUMBER OF SEQ ID NOS: 362830
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8133
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-11-136-527-8133

Alignment Scores:
Pred. No.:                         Length:           600
Score:                  931        Matches:          8
Score:                  8.00
```

```
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 837176
;     LENGTH: 602
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-925-065A-837176

Alignment Scores:
Pred. No.:                            934          Length:              602
Score:                               8.00          Matches:               8
Percent Similarity:                 100.0%         Conservative:          0
Best Local Similarity:              100.0%         Mismatches:            0
Query Match:                          3.8%         Indels:                0
DB:                                    6           Gaps:                  0

US-09-989-890-238 (1-212) x US-09-925-065A-837176 (1-602)

Qy         25 ProSerLeuAlaLeuLeuSer 32
              ||||||||||||||||||||||
Db        198 CCCAGCTCTTTAGCTCTCTTGAGT 175


RESULT 71
US-09-925-065A-571840/c
; Sequence 571840, Application US/09925065A
; Publication No. US2004018104 8A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 571840
;     LENGTH: 603
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-925-065A-571840

Alignment Scores:
Pred. No.:                            935          Length:              603
Score:                               8.00          Matches:               8
Percent Similarity:                 100.0%         Conservative:          0
Best Local Similarity:              100.0%         Mismatches:            0
Query Match:                          3.8%         Indels:                0
DB:                                    6           Gaps:                  0

US-09-989-890-238 (1-212) x US-09-925-065A-571840 (1-603)

Qy         28 LeuAlaLeuSerArgProLeu 35
              |||||||||||||||||||||
Db        552 CTGGCTTCCTCCTCCAGACCACTC 529


RESULT 72
US-09-925-065A-630748
; Sequence 630748, Application US/09925065A
; Publication No. US2004018104 8A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
```

```
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 630748
;     LENGTH: 609
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-925-065A-630748

Alignment Scores:
Pred. No.:                            943          Length:              609
Score:                               8.00          Matches:               8
Percent Similarity:                 100.0%         Conservative:          0
Best Local Similarity:              100.0%         Mismatches:            0
Query Match:                          3.8%         Indels:                0
DB:                                    6           Gaps:                  0

US-09-989-890-238 (1-212) x US-09-925-065A-630748 (1-609)

Qy         25 ProSerSerLeuAlaLeuLeuSer 32
              ||||||||||||||||||||||||
Db        166 CCATCCAGCTTGGCTCTGCTGTCG 189


RESULT 73
US-09-925-065A-893900/c
; Sequence 893900, Application US/09925065A
; Publication No. US2004018104 8A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 893900
;     LENGTH: 611
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-925-065A-893900

Alignment Scores:
Pred. No.:                            946          Length:              611
Score:                               8.00          Matches:               8
Percent Similarity:                 100.0%         Conservative:          0
Best Local Similarity:              100.0%         Mismatches:            0
Query Match:                          3.8%         Indels:                0
DB:                                    6           Gaps:                  0
```

US-09-989-890-238 (1-212) x US-09-925-065A-893900 (1-611)

```
Qy        24 LeuProSerSerLeuAlaLeuLeu 31
             ||||||||||||||||||||||||
Db       506 TTGCCTAGCAGCCTAGCACTTTTG 483
```

RESULT 74
US-09-925-065A-894636/c
; Sequence 894636, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 894636
;   LENGTH: 621
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-894636

Alignment Scores:
Pred. No.:   960   Length: 621
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0
```

US-09-989-890-238 (1-212) x US-09-925-065A-894636 (1-621)

```
Qy        24 LeuProSerSerLeuAlaLeuLeu 31
             ||||||||||||||||||||||||
Db       101 TTGCCTAGCAGCCTAGCACTTTTG 78
```

RESULT 75
US-09-925-065A-464419/c
; Sequence 464419, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 464419
;   LENGTH: 622
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-464419

Alignment Scores:
Pred. No.:   961   Length: 622
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0
```

US-09-989-890-238 (1-212) x US-09-925-065A-464419 (1-622)

```
Qy        27 SerLeuAlaLeuLeuSerArgPro 34
             ||||||||||||||||||||||||
Db       287 AGTCCGCCTCGTTGTCCAGGCCG 264
```

RESULT 76
US-09-925-065A-672678/c
; Sequence 672678, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 672678
;   LENGTH: 623
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-672678

Alignment Scores:
Pred. No.:   963   Length: 623
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0
```

US-09-989-890-238 (1-212) x US-09-925-065A-672678 (1-623)

```
Qy        54 LeuProSerAlaSerAlaAlaAla 61
             ||||||||||||||||||||||||
Db       589 CTTCCCTCTGCTTCAGCAGCTGCT 566
```

RESULT 77
US-09-925-065A-686260/c
; Sequence 686260, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A

; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 686260
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-686260

Alignment Scores:
Pred. No.:   963   Length: 623
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-686260 (1-623)

Qy   26 SerSerLeuAlaLeuLeuSerArg 33
         ||||||||||||||||||||||||
Db   483 TCGAGTCTTGCTCTGTTGTCCAGG 460

RESULT 78
US-09-925-065A-31084/c
; Sequence 31084, Application US/09925065A
; Publication No. US2004018048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31084
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-31084

Alignment Scores:
Pred. No.:   964   Length: 624
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-31084 (1-624)

Qy   27 SerLeuAlaLeuLeuSerArgPro 34

Db   356 AGTCTTGCTCTGTTGTCCAGGCCA 333

RESULT 79
US-09-925-065A-749727/c
; Sequence 749727, Application US/09925065A
; Publication No. US2004018048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 749727
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-749727

Alignment Scores:
Pred. No.:   964   Length: 624
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-749727 (1-624)

Qy   53 GlyLeuProSerAlaSerAlaAla 60
         ||||||||||||||||||||||||
Db   573 GGGCTCCATCTGCCTCTGCTGCT 550

RESULT 80
US-09-925-065A-822330/c
; Sequence 822330, Application US/09925065A
; Publication No. US2004018048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 822330
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-925-065A-822330

Alignment Scores:
Pred. No.:
Score: 964
Percent Similarity: 8.00   Length: 624
Best Local Similarity: 100.0%   Matches: 8
Query Match: 100.0%   Conservative: 0
DB: 3.8%   Mismatches: 0
                           6   Indels: 0
                                                   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-822330 (1-624)

Qy   53 GlyLeuProSerAlaSerAlaAla 60
             ||||||||||||||||||||||||
Db   573 GGGCTCCCATCTGCCTCTGCTGCT 550

RESULT 81
US-09-925-065A-883750
; Sequence 883750, Application US/09925065A
; Publication No. US2004018104A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 883750
;   LENGTH: 637
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-883750

Alignment Scores:
Pred. No.:
Score: 982
Percent Similarity: 8.00   Length: 637
Best Local Similarity: 100.0%   Matches: 8
Query Match: 100.0%   Conservative: 0
DB: 3.8%   Mismatches: 0
                           6   Indels: 0
                                                   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-883750 (1-637)

Qy   28 LeuAlaLeuSerArgProLeu 35
             |||||||||||||||||||||
Db   243 TTAGCACTCCTTTCAAGACCTCTC 266

RESULT 82
US-09-925-065A-910175
; Sequence 910175, Application US/09925065A
; Publication No. US2004018104A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147

;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 910175
;   LENGTH: 637
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-910175

Alignment Scores:
Pred. No.:
Score: 982
Percent Similarity: 8.00   Length: 637
Best Local Similarity: 100.0%   Matches: 8
Query Match: 100.0%   Conservative: 0
DB: 3.8%   Mismatches: 0
                           6   Indels: 0
                                                   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-910175 (1-637)

Qy   28 LeuAlaLeuSerArgProLeu 35
             |||||||||||||||||||||
Db   243 TTAGCACTCCTTTCAAGACCTCTC 266

RESULT 83
US-09-925-065A-709912/c
; Sequence 709912, Application US/09925065A
; Publication No. US2004018104A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 709912
;   LENGTH: 639
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-709912

Alignment Scores:
Pred. No.:
Score: 985
Percent Similarity: 8.00   Length: 639
Best Local Similarity: 100.0%   Matches: 8
Query Match: 100.0%   Conservative: 0
DB: 3.8%   Mismatches: 0
                           6   Indels: 0
                                                   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-709912 (1-639)

Qy   34 ProLeuSerProProProAlaAla 41
             ||||||||||||||||||||||||
Db   483 CCTCTCCTCCTCCTCCTCCTGCTGCT 460

RESULT 84

```
US-09-925-065A-441386/c
; Sequence 441386, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 441386
;   LENGTH: 662
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-441386

Alignment Scores:
Pred. No.:                1.02e+03     Length:            662
Score:                    8.00         Matches:           8
Percent Similarity:       100.0%       Conservative:      0
Best Local Similarity:    100.0%       Mismatches:        0
Query Match:              3.8%         Indels:            0
DB:                       6            Gaps:              0

US-09-989-890-238 (1-212) x US-09-925-065A-441386 (1-662)

Qy       55 ProSerAlaSerAlaAlaAlaGly 62
            ||||||||||||||||||||
Db      280 CCATCTGCCTCTGCTGCGGCAGT 257


RESULT 85
US-09-925-065A-4577/c
; Sequence 4577, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4577
;   LENGTH: 664
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-4577

Alignment Scores:
Pred. No.:                1.02e+03     Length:            664
Score:                    8.00         Matches:           8
Percent Similarity:       100.0%       Conservative:      0
Best Local Similarity:    100.0%       Mismatches:        0
Query Match:              3.8%         Indels:            0
DB:                       6            Gaps:              0

US-09-989-890-238 (1-212) x US-09-925-065A-4577 (1-664)

Qy       24 LeuProSerLeuAlaLeuLeu 31
            |||||||||||||||||||||
Db       85 CTACCTTCTTCACTAGCCTTGCTG 62


RESULT 86
US-09-925-065A-4578/c
; Sequence 4578, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4578
;   LENGTH: 664
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-4578

Alignment Scores:
Pred. No.:                1.02e+03     Length:            664
Score:                    8.00         Matches:           8
Percent Similarity:       100.0%       Conservative:      0
Best Local Similarity:    100.0%       Mismatches:        0
Query Match:              3.8%         Indels:            0
DB:                       6            Gaps:              0

US-09-989-890-238 (1-212) x US-09-925-065A-4578 (1-664)

Qy       24 LeuProSerLeuAlaLeuLeu 31
            |||||||||||||||||||||
Db       85 CTACCTTCTTCACTAGCCTTGCTG 62


RESULT 87
US-09-925-065A-4579/c
; Sequence 4579, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
```

```
  ;    PRIOR FILING DATE: 2001-01-16
  ;    PRIOR APPLICATION NUMBER: US 60/289,846
  ;    PRIOR FILING DATE: 2001-05-09
  ;    NUMBER OF SEQ ID NOS: 957086
  ;    SOFTWARE: FastSEQ for Windows Version 4.0
  ;    SEQ ID NO 4579
  ;    LENGTH: 664
  ;    TYPE: DNA
  ;    ORGANISM: Homo sapiens
  US-09-925-065A-4579

  Alignment Scores:
  Pred. No.:                1.02e+03          Length:            664
  Score:                    8.00              Matches:           8
  Percent Similarity:       100.0%            Conservative:      0
  Best Local Similarity:    100.0%            Mismatches:        0
  Query Match:              3.8%              Indels:            0
  DB:                                         Gaps:              6

  US-09-989-890-238 (1-212) x US-09-925-065A-4579 (1-664)

  Qy      24 LeuProSerSerLeuAlaLeuLeu 31
             ||||||||||||||||||||||||
  Db      85 CTACCTTCTTCACTAGCCTTGCTG 62

  RESULT 88
  US-11-103-957-91/c
  ; Sequence 91, Application US/11103957
  ; Publication No. US20050281847A1
  ; GENERAL INFORMATION:
  ;    APPLICANT: Berthet, Francois-Xavier Jacques
  ;    APPLICANT: Lobet, Yves
  ;    APPLICANT: Poolman, Jan
  ;    APPLICANT: Verlant, Vincent Georges Christian Louis
  ;    TITLE OF INVENTION: Vaccine Composition
  ;    FILE REFERENCE: B45261
  ;    CURRENT APPLICATION NUMBER: US/11/103,957
  ;    CURRENT FILING DATE: 2005-04-12
  ;    PRIOR APPLICATION NUMBER: US/10/467,534
  ;    PRIOR FILING DATE: 2004-02-03
  ;    PRIOR APPLICATION NUMBER: PCT/EP02/01356
  ;    PRIOR FILING DATE: 2002-02-08
  ;    PRIOR APPLICATION NUMBER: GB 0103169.9
  ;    PRIOR FILING DATE: 2001-02-08
  ;    NUMBER OF SEQ ID NOS: 108
  ;    SOFTWARE: FastSEQ for Windows Version 4.0
  ;    SEQ ID NO 91
  ;    LENGTH: 710
  ;    TYPE: DNA
  ;    ORGANISM: Neisseria gonorrhoeae
  US-11-103-957-91

  Alignment Scores:
  Pred. No.:                1.08e+03          Length:            710
  Score:                    8.00              Matches:           8
  Percent Similarity:       100.0%            Conservative:      0
  Best Local Similarity:    100.0%            Mismatches:        0
  Query Match:              3.8%              Indels:            0
  DB:                                         Gaps:              12

  US-09-989-890-238 (1-212) x US-11-103-957-91 (1-710)

  Qy      55 ProSerAlaSerAlaAlaAlaGly 62
             ||||||||||||||||||||||||
  Db     202 CCTTCCGCCAGTGCGGCGCCGGG 179

  RESULT 89
  US-11-077-619-41/c
  ; Sequence 41, Application US/11077619
  ; Publication No. US20060040279A1
  ; GENERAL INFORMATION:
  ;    APPLICANT: Feesche, Joerg
  ;    APPLICANT: Maurer, Karl-Heinz
```

```
  ;    APPLICANT: Breves, Roland
  ;    APPLICANT: Schweder, Thomas
  ;    APPLICANT: Hecker, Michael
  ;    APPLICANT: Juergen, Britta
  ;    APPLICANT: Voigt, Birgit
  ;    TITLE OF INVENTION: DNA CHIPS USED FOR BIOPROCESS CONTROL
  ;    FILE REFERENCE: HENK-0122 / H5692
  ;    CURRENT APPLICATION NUMBER: US/11/077,619
  ;    CURRENT FILING DATE: 2005-03-11
  ;    PRIOR APPLICATION NUMBER: PCT/EP2003/009979
  ;    PRIOR FILING DATE: 2003-09-09
  ;    PRIOR APPLICATION NUMBER: DE 10242433.0
  ;    PRIOR FILING DATE: 2002-09-11
  ;    NUMBER OF SEQ ID NOS: 130
  ;    SOFTWARE: PatentIn version 3.3
  ;    SEQ ID NO 41
  ;    LENGTH: 829
  ;    TYPE: DNA
  ;    ORGANISM: Escherichia coli
  ;    FEATURE:
  ;    NAME/KEY: gene
  ;    LOCATION: (1)..(829)
  ;    OTHER INFORMATION: ibpB
  ;    FEATURE:
  ;    NAME/KEY: CDS
  ;    LOCATION: (201)..(629)
  US-11-077-619-41

  Alignment Scores:
  Pred. No.:                1.24e+03          Length:            829
  Score:                    8.00              Matches:           8
  Percent Similarity:       100.0%            Conservative:      0
  Best Local Similarity:    100.0%            Mismatches:        0
  Query Match:              3.8%              Indels:            0
  DB:                                         Gaps:              9

  US-09-989-890-238 (1-212) x US-11-077-619-41 (1-829)

  Qy      75 AlaAlaProAlaCysLeuLeuArg 82
             ||||||||||||||||||||||||
  Db     429 GCTGCTCCGGCGTGCCTTTTTACGC 406

  RESULT 90
  US-10-821-234-780/c
  ; Sequence 780, Application US/10821234
  ; Publication No. US2005025511A1
  ; GENERAL INFORMATION:
  ;    APPLICANT: Stache-Crain, Birgit
  ;    APPLICANT: Labat, Ivan
  ;    APPLICANT: Andarmani, Susan
  ;    APPLICANT: Tang, Y. Tom
  ;    TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
  ;    FILE REFERENCE: 821A
  ;    CURRENT APPLICATION NUMBER: US/10/821,234
  ;    CURRENT FILING DATE: 2004-04-07
  ;    PRIOR APPLICATION NUMBER: US 60/462,047
  ;    PRIOR FILING DATE: 2003-04-07
  ;    NUMBER OF SEQ ID NOS: 1704
  ;    SOFTWARE: pt_SEQ_genes Version 1.0
  ;    SEQ ID NO 780
  ;    LENGTH: 855
  ;    TYPE: DNA
  ;    ORGANISM: Homo sapiens
  US-10-821-234-780

  Alignment Scores:
  Pred. No.:                1.28e+03          Length:            855
  Score:                    8.00              Matches:           8
  Percent Similarity:       100.0%            Conservative:      0
  Best Local Similarity:    100.0%            Mismatches:        0
  Query Match:              3.8%              Indels:            0
  DB:                                         Gaps:              0
```

US-09-989-890-238 (1-212) x US-10-821-234-780 (1-855)

```
Qy      37  ProProProAlaAlaCysSerGly  44
            ||||||||||||||||||||||
Db     124  CCTCCTCCAGCTGCTTGCCAGCGGT  101
```

RESULT 91
US-11-128-061-61
; Sequence 61, Application US/1128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
;   APPLICANT: Melville, Mark W.
;   APPLICANT: Charlebois, Timothy S.
;   APPLICANT: Mounts, William M.
;   APPLICANT: Hann, Louane E.
;   APPLICANT: Sinacore, Martin S.
;   APPLICANT: Leonard, Mark W.
;   APPLICANT: Brown, Eugene L.
;   APPLICANT: Miller, Christopher P.
;   TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO MONITOR GENE EXPRESSION
;   FILE REFERENCE: 01997.027701
;   CURRENT APPLICATION NUMBER: US/11/128,061
;   CURRENT FILING DATE: 2005-05-11
;   PRIOR APPLICATION NUMBER: US 60/570,425
;   PRIOR FILING DATE: 2004-05-11
;   NUMBER OF SEQ ID NOS: 7285
;   SOFTWARE: PatentIn version 3.3
; SEQ ID NO 61
;   LENGTH: 1002
;   TYPE: DNA
;   ORGANISM: Cricetulus griseus
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (492)..(512)
;   OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-61

Alignment Scores:
Pred. No.:   1.47e+03   Length: 1002
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-128-061-61 (1-1002)

```
Qy      23  AlaLeuProSerLeuAlaLeu  30
            ||||||||||||||||||||
Db     210  GCCCTGCCCAGTTCTCTGGCCCTA  233
```

RESULT 92
US-11-128-049-61
; Sequence 61, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
;   APPLICANT: Melville, Mark W.
;   APPLICANT: Charlebois, Timothy S.
;   APPLICANT: Mounts, William M.
;   APPLICANT: Hann, Louane E.
;   APPLICANT: Sinacore, Martin S.
;   APPLICANT: Leonard, Mark W.
;   APPLICANT: Brown, Eugene L.
;   APPLICANT: Miller, Christopher P.
;   TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
;   TITLE OF INVENTION: MAKING AND USING SAME
;   FILE REFERENCE: 01997.027700
;   CURRENT APPLICATION NUMBER: US/11/128,049
;   CURRENT FILING DATE: 2005-05-11
;   PRIOR APPLICATION NUMBER: US 60/570,425
;   PRIOR FILING DATE: 2004-05-11
;   NUMBER OF SEQ ID NOS: 7285
;   SOFTWARE: PatentIn version 3.3
; SEQ ID NO 61
;   LENGTH: 1002
;   TYPE: DNA
;   ORGANISM: Cricetulus griseus
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (492)..(512)
;   OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-61

Alignment Scores:
Pred. No.:   1.47e+03   Length: 1002
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-128-049-61 (1-1002)

```
Qy      23  AlaLeuProSerLeuAlaLeu  30
            ||||||||||||||||||||
Db     210  GCCCTGCCCAGTTCTCTGGCCCTA  233
```

RESULT 93
US-09-925-065A-63837/c
; Sequence 63837, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63837
;   LENGTH: 1038
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-63837

Alignment Scores:
Pred. No.:   1.52e+03   Length: 1038
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-63837 (1-1038)

```
Qy      27  SerLeuAlaLeuSerArgPro  34
            ||||||||||||||||||||
Db     768  AGTCTTGCTCGTTGTCCAGGCCA  745
```

RESULT 94
US-09-925-065A-63838/c
; Sequence 63838, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,096
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63838
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-63838

Alignment Scores:
Pred. No.:   1.52e+03   Length: 1038
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 6   Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-63838 (1-1038)

Qy   27 SerLeuAlaLeuSerArgPro 34
            |||||||||||||||||||||
Db   768 AGTCTTGCTCTGTTGTCCAGGCCA 745

RESULT 95
US-09-925-065A-63839/c
; Sequence 63839, Application US2004018104841
; Publication No. US2004018104841
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63839
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-63839

Alignment Scores:
Pred. No.:   1.52e+03   Length: 1038
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: Gaps: 0

US-09-989-890-238 (1-212) x US-09-925-065A-63839 (1-1038)

Qy   27 SerLeuAlaLeuSerArgPro 34
            |||||||||||||||||||||
Db   768 AGTCTTGCTCTGTTGTCCAGGCCA 745

RESULT 96
US-11-091-883-130/c
; Sequence 130, Application US/11091883
; Publication No. US20060024693A1
; GENERAL INFORMATION:
; APPLICANT: CIBELLI, JOSE
; APPLICANT: FERNANDEZ, EMILIO O.
; APPLICANT: JORDAO DE MEGALHAES, GUILHERME
; APPLICANT: KOCABAS, ARIF
; APPLICANT: CROSBY, JAVIER A.
; TITLE OF INVENTION: IDENTIFICATION OF GENES OR POLYPEPTIDES THE EXPRESSION OF WHICH
; TITLE OF INVENTION: CORRELATES TO FERTILITY, OVARIAN FUNCTION AND/OR FETAL/NEWBORN
; TITLE OF INVENTION: VIABILITY
; FILE REFERENCE: 53942US
; CURRENT APPLICATION NUMBER: US/11/091,883
; CURRENT FILING DATE: 2005-03-29
; PRIOR APPLICATION NUMBER: 60/556,875
; PRIOR FILING DATE: 2004-03-29
; NUMBER OF SEQ ID NOS: 513
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 130
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-091-883-130

Alignment Scores:
Pred. No.:   1.53e+03   Length: 1044
Score: 8.00   Matches: 8
Percent Similarity: 100.0%   Conservative: 0
Best Local Similarity: 100.0%   Mismatches: 0
Query Match: 3.8%   Indels: 0
DB: 12   Gaps: 0

US-09-989-890-238 (1-212) x US-11-091-883-130 (1-1044)

Qy   37 ProProProAlaAlaCysSerGly 44
            |||||||||||||||||||||
Db   242 CCTCCTCCAGCTGCTTGCAGCGGT 219

RESULT 97
US-09-925-065A-576
; Sequence 576, Application US/09925065A
; Publication No. US2004018104841
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 576
; LENGTH: 1056

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-576

Alignment Scores:
Pred. No.:                              Length:         1056
Score:                  1.54e+03        Matches:        8
Percent Similarity:     100.0%          Conservative:   0
Best Local Similarity:  100.0%          Mismatches:     0
Query Match:            3.8%            Indels:         0
DB:                     6               Gaps:           0

US-09-989-890-238 (1-212) x US-09-925-065A-576 (1-1056)

Qy    27 SerLeuAlaLeuSerArgPro 34
         ||||||||||||||||||||
Db   290 AGTCTTGCTCTGTTGTCCAGGCCA 313

RESULT 98
US-09-925-065A-577
;   Sequence 577, Application US/09925065A
;   Publication No. US20040181048A1
;   GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 577
;   LENGTH: 1056
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-577

Alignment Scores:
Pred. No.:                              Length:         1056
Score:                  1.54e+03        Matches:        8
Percent Similarity:     100.0%          Conservative:   0
Best Local Similarity:  100.0%          Mismatches:     0
Query Match:            3.8%            Indels:         0
DB:                     6               Gaps:           0

US-09-989-890-238 (1-212) x US-09-925-065A-577 (1-1056)

Qy    27 SerLeuAlaLeuSerArgPro 34
         ||||||||||||||||||||
Db   290 AGTCTTGCTCTGTTGTCCAGGCCA 313

RESULT 99
US-09-925-065A-578
;   Sequence 578, Application US/09925065A
;   Publication No. US20040181048A1
;   GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
```

```
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 578
;   LENGTH: 1056
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-578

Alignment Scores:
Pred. No.:                              Length:         1056
Score:                  1.54e+03        Matches:        8
Percent Similarity:     100.0%          Conservative:   0
Best Local Similarity:  100.0%          Mismatches:     0
Query Match:            3.8%            Indels:         0
DB:                     6               Gaps:           0

US-09-989-890-238 (1-212) x US-09-925-065A-578 (1-1056)

Qy    27 SerLeuAlaLeuSerArgPro 34
         ||||||||||||||||||||
Db   290 AGTCTTGCTCTGTTGTCCAGGCCA 313

RESULT 100
US-09-925-065A-579
;   Sequence 579, Application US/09925065A
;   Publication No. US20040181048A1
;   GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single
;   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.135
;   CURRENT APPLICATION NUMBER: US/09/925,065A
;   CURRENT FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: US 60/243,096
;   PRIOR FILING DATE: 2000-10-24
;   PRIOR APPLICATION NUMBER: US 60/252,147
;   PRIOR FILING DATE: 2000-11-20
;   PRIOR APPLICATION NUMBER: US 60/250,092
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: US 60/261,766
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/289,846
;   PRIOR FILING DATE: 2001-05-09
;   NUMBER OF SEQ ID NOS: 957086
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 579
;   LENGTH: 1056
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-579

Alignment Scores:
Pred. No.:                              Length:         1056
Score:                  1.54e+03        Matches:        8
Percent Similarity:     100.0%          Conservative:   0
Best Local Similarity:  100.0%          Mismatches:     0
Query Match:            3.8%            Indels:         0
DB:                     6               Gaps:           0

US-09-989-890-238 (1-212) x US-09-925-065A-579 (1-1056)

Qy    27 SerLeuAlaLeuSerArgPro 34
         ||||||||||||||||||||
Db   290 AGTCTTGCTCTGTTGTCCAGGCCA 313
```

us-09-989-890-238.oligo_p2n.rnpbn

Fri Mar 17 09:43:20 2006

Search completed: March 17, 2006, 07:00:40
Job time : 1316 secs